Structural Dynamics and Allosteric Signaling in Ionotropic Glutamate Receptors

by

Anindita Dutta

B.E Birla Institute of Technology and Science, Pilani, 2008

M.Sc Birla Institute of Technology and Science, Pilani, 2008

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

School of Medicine

This dissertation was presented

by

Anindita Dutta

It was defended on

October 24th, 2013

and approved by

Dr. Xiang-Qun (Sean) Xie, Professor, Pharmaceutical Sciences, University of Pittsburgh

Dr. Michael Grabe, Associate Professor, Department of Biological Sciences, University of Pittsburgh

Dr. Jamie Carbonell, Professor, School of Computer Science (LTI), Carnegie Mellon University

Dr. Ingo Greger, Group Leader, MRC Laboratory of Molecular Biology, Cambridge, UK

Dissertation Advisor: Dr. Ivet Bahar, Professor, Department of Computational and Systems Biology, University of Pittsburgh

Structural Dynamics and Allosteric Signaling in Ionotropic Glutamate Receptors

Anindita Dutta, PhD

University of Pittsburgh, 2013

Copyright © by Anindita Dutta

2013

Ionotropic glutamate receptors (iGluRs) are ligand-gated ion channels that mediate excitatory neurotransmission events in the central nervous system. All distinct classes of iGluRs (AMPA, NMDA, Kainate) are composed of an N-terminal domain (NTD) and a ligand-binding domain (LBD) in their extracellular domain, a transmembrane domain (TMD) and an intracellular carboxy-terminal domain (CTD). Ligand binding to the LBD facilitates ion channel activation. The NTDs modulate channel gating allosterically in NMDA receptors (NMDARs). A similar function of the NTD in AMPA receptors (AMPARs) is still a matter of debate. Taking advantage of recently resolved structures of the NTD and the intact AMPAR, the main focus of this dissertation is a comprehensive examination of iGluR NTD structural dynamics, ligand binding and allosteric potential of AMPARs. We use a multiscale, multi-dimensional approach using coarse-grained network models and all-atom simulations for structural analyses and information theoretic approaches for examination of evolutionary correlations. Our major contribution has been the characterization of the global motions favored by iGluR NTD architecture. These intrinsic motions favor ligand binding in NMDAR NTDs and are also shared by other iGluR NTDs. We also identified structural determinants of flexibility in AMPARs and confirmed their role through *in silico* mutants.

The overall similarity in collective dynamics among iGluRs hints at a putative allosteric capacity of non-NMDARs and has propelled the elucidation of interdomain and intersubunit coupling in the intact AMPAR. To this end, we identified “effector” and “sensor” regions in AMPARs using a perturbation-response technique. We identified potentially functional residues that enable information propagation between effector regions and proposed an efficient mechanism of allosteric communication based on a combination of tools including network models, graph theoretical methods and sequence analyses.

Finally, we assessed the "druggability" of iGluR NTDs using molecular dynamics simulations in the presence of probe molecules containing fragments shared by drug-like molecules. Based on our study, we offer key insights into the ligand-binding landscape of iGluR NTD monomers and dimers, and we also identify a novel ligand-binding site in AMPAR dimers. These findings open an avenue of searching for molecules able to bind to iGluR NTDs and allosterically modulate receptor activity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to sincerely thank my adviser Prof. Ivet Bahar, for being an amazing mentor, whose guidance and strong support, both personally and professionally, has made my graduate student life a smooth sailing and memorable experience. She has been instrumental in fostering independent thought, clarity and focus in ideas, and instigating new dimensions in research through a very positive and encouraging work atmosphere. I feel privileged to have had her as my guide, and she will always be an inspiration to me in all my future endeavors.

I would also like to thank my collaborator, Prof. Ingo Greger, whose crisp biological insights have made a strong impact in shaping my thoughts and research ideas. I am also grateful to my other thesis committee members, Dr. Michael Grabe, Dr. Jaime Carbonell, Dr. Sean Xie, for their insightful comments and critiques.

I would also like to take the opportunity to thank Dr. Indira Shrivastava, for being a patient and very encouraging colleague and co-author and I am really glad that I had an opportunity to work closely with her. A very special thanks to Dr. Ahmet Bakan, who had the confidence to let me be a contributor to his flagship *ProDy* project, urging me to learn and excel in new skills. I also feel fortunate to have had extremely helpful lab members (present and ex), Drs. Ying Liu, Lidio Meireles, Zheng Yang, Mert Gur, Timothy Lezon, Elia Zomot, Ignacio General, Filippo Pullara, Mary Chang, Kristina Paris and fellow graduate students Murat Can,

Mao, Kaitlyn and Chang. I have had a great time working alongside them, and having many a fruitful discussions about science and life in general.

I feel lucky to have been accepted to the Joint CMU-PITT program, and I am happy to have met and interacted with all the students and faculty of the program. In the department, I am grateful to Kelly, Sandy and Nancy for always being there to help.

A special acknowledgement to fellow BST3 friends Grace, Ginny, Andrej and John without whose company graduate life would be much less enjoyable. I will always cherish the time spent in the final year of thesis with Grace and Ginny that made work seem like fun.

I would like to thank my friends in Pittsburgh, who have made my stay there a memorable experience, especially Rachana, Jineta and Sourish, whose friendship has meant so much to me and will continue to do so.

Lastly, I would like to thank my family, my sister Paromita, my father Mr. Prabuddha Dutta and my grandmother for always being so supportive in whatever I do. And finally I would like to dedicate my thesis to my mother, Mrs. Madhumita Dutta, who couldn't see me graduate, but I hope she is very proud of me.

1.0 INTRODUCTION

Most proteins are molecular machines. They achieve their function because they possess the ability to undergo the structural changes required for their function. These structural changes may vary over a broad range of scales, from atomic fluctuations and side chain rotations, to collective domain or subunit movements. Yet, even the large-scale movements do not, in principle, alter the native 'fold'; the packing of secondary structural elements and/or the distribution of tertiary and quaternary contacts between residues remain unchanged for the most part. The collective motions allow for cooperative rearrangements of domains with respect to each other, or loop motions, which maintain the overall architecture/fold, or allow the protein to restore its original conformation, similar to an elastic material. Here cooperative motions refer to conformational changes that collectively engage large portions of the structure and usually lie at the lowest frequency end of the spectrum of modes accessible to the protein under physiological conditions. Not surprisingly, normal mode analysis (NMA) of elastic network models (ENMs) of protein structural coordinates have been exploited in unraveling protein dynamics, based on the premise that proteins have 'intrinsic' abilities, uniquely encoded by their 3-dimensional fold, to sample functional fluctuations near their native state (Bahar et al., 2007; Bahar et al., 2010a).

From a statistical thermodynamics perspective, the accessible conformations result from the equilibrium fluctuations near the global free energy minimum. Some directions of fluctuations are more probable than others. These are the 'soft modes' of motions; they evolve

along a direction, on the multidimensional energy landscape, which by definition involves a relatively small energy ascent for a given deformation. ENM-NMAs a well as simplified models such as the Gaussian Network Model (GNM), (Bahar et al., 1997; Haliloglu et al., 1997) rooted in the statistical thermodynamics of polymer networks, (Flory, 1976) have been broadly used in the last decade to extract these soft modes of motion. These same modes are also most likely to be frequently recruited to achieve function, following the hypothesis that structures may have evolved to access most easily the most functional movements. A large body of work has shown that in fact these modes conform to the structural changes associated with ligand binding and allosteric transitions showing their relevance to function (Bahar et al., 2010a; Bahar et al., 2010b). Chapter 2 provides an overview of the theory and assumptions underlying these approaches.

The practical utility of ENM-NMA of proteins is the predictions of collective changes of structures that are expected to be most readily accessible, or to generate alternative conformations that satisfy certain requirements, e.g. optimal binding of a substrate (Bakan and Bahar, 2009; Cavasotto et al., 2005; Tobi and Bahar, 2005). Modifications of ENM-NMA can be used to assess the effect of environment modeled as a perturbation to the system (Ming and Wall, 2005; Zheng and Brooks, 2005), and to study the effect of residue perturbation modeled as forces to individual network nodes (Atilgan and Atilgan, 2009). Also biomolecular structures need to be optimally 'wired' to efficiently transmit signals, to couple or exchange their chemical and mechanical energies, or induce allosteric responses. ENM representation of proteins permits us to utilize graph theory, along with information theoretic approaches to identify the sites that play a critical role in the flow of information across the 'network' of amino acids in the folded state. The collective fluctuations predicted by the GNM also define signal propagation mechanisms, as

derived using a Markovian model in a recent study (Chennubhotla and Bahar, 2007). Certain 'nodes' are distinguished by their effective signal transmission, or allosteric communication, properties. Examination of the identity of such key sites reveals a striking preponderance of active sites (e.g., catalytic residues, metal-binding sites), suggesting that efficient transmission of signals is an inherent, structure-encoded property of active sites. Conceivably, active sites of proteins evolved to be positioned at sites that lend themselves to efficient communication (Chennubhotla and Bahar, 2007; Yang and Bahar, 2005). A specific application to metal-binding sites is highlighted in Chapter 2.

One clear limitation of this method is its applicability to near-native conditions only, because it implicitly assumes linear dynamics, all interactions being approximated by harmonic potentials. This assumption is strictly valid in the immediate neighborhood of the global energy minimum. Thus NMA is unable to capture anharmonic effects and non-linear dynamics as well as solvent damping effects that proteins are subjected to (Ma, 2005). So coarse-graining presents the advantage of eliminating certain degrees of freedom and resulting in a 'smoother' energy landscape where local minima are overlooked (or overcome) during collective motions providing access to substates separated by low energy barriers, but this comes at the cost of losing atomic level accuracy. The results from NMA should therefore be interpreted in the context of these approximations.

For applications like studying specific interactions in protein-protein or protein-ligand interactions, it becomes important for modeling proteins at all-atom level. Molecular dynamics (MD) is a simulation tool widely used to capture the time evolution of physical motions of interacting (modeled using molecular mechanics force-fields) protein atoms that incorporates all molecular driving forces. The use of all-atom models for MD runs in the presence of only water

(Chapter 3) and water and organic probe molecules (Chapter 5) enables the study of molecular phenomena and provides information about side chain specific local interactions. MD simulation based exploration of putative "druggable" sites enable the incorporation of conformational flexibility of the proteins and drug-like probe molecules and provide more realistic detection of binding sites and the maximal binding free energy (Bakan et al., 2012). The structure and dynamics that govern the binding propensities of proteins are explored in Chapter 5. Thus we take a multi-scale approach using both coarse-grained and all-atom models to decipher structural dynamics of proteins.

Analysis of homologous sequences of protein families can help uncover residues that are under evolutionary constraints due to their importance in structure and function of the protein. At the same time characterization of correlations existing between sequence evolution and intrinsic structural dynamics is of utmost importance towards understanding the molecular mechanisms of function and how such functionalities may have evolved over time and species. There are a number of independent software packages that allow for exploration of phylogenetic data (Ashkenazy et al., 2010; Eyal et al., 2007; Morgan et al., 2006; Wainreb et al., 2011) or structural protein dynamics (Eyal et al., 2006; Suhre and Sanejouand, 2004), but a comprehensive software suite enabling fast and efficient comparative analysis of evolutionary patterns and conformations dynamics is still lacking. To meet this goal, we developed an extension of the open source python package *ProDy (*Bakan et al. 2011), originally developed for inferring information on *pro*tein *dyn*amics from ensembles of structural data. We have made significant additions that permit access to databases that carry sequence alignment data, allow for intuitive and competent handling and refinement of large sequence alignment files, and enable users to evaluate conservation and coevolution propensities using information theoretic

approaches, and analyze them in relation to intrinsic structural dynamics of the examined family of proteins inferred from experiments and computations. Thus, we take a multi-dimensional approach facilitating analysis of protein family members capturing and uniting knowledge from structural and evolutionary dynamics.

Most of our study using the methods described above is focused on ionotropic glutamate receptors (iGluRs). These receptors are ligand-gated ion channels that mediate the majority of the excitatory neurotransmission in the central nervous system. They are important drug targets because of their implication in neurodegenerative diseases and neurological disorders. Exploring the structural basis of their function and ligand-binding properties and identification of new functional sites offers a prolific avenue of research towards developing improved CNS therapeutics. The following section describes the iGluRs.

1.1 IONOTROPIC GLUTAMATE RECEPTOR FAMILY

Synaptic transmission between neurons is mediated by the release of neurotransmitters from pre-synaptic axon terminals, and by their subsequent binding to receptors in the post-synaptic cells for signal propagation. The majority of fast excitatory neurotransmission in the vertebrate central nervous system is mediated by glutamate receptors (GluRs), which are classified as ionotropic or metabotropic. Ionotropic glutamate receptors (iGluRs) are tetrameric ligand-gated ion channels that stimulate membrane depolarization of the post-synaptic cell by selective translocation of ions (K^+, Na^+, Ca^{2+}) upon ligand binding (Madden, 2002). iGluR activity is important to physiological processes like learning, memory and modulation of neuronal excitability (Bigge, 1999; Dingledine et al., 1999). These receptors have also been implicated in pathological

conditions like chronic pain, ischemia (Dingledine et al., 1999), and in neurological disorders like Parkinson's and Huntington's disease, thereby sparking pharmaceutical interest as potential drug targets (Bowie, 2008).

The pharmacologically distinct classes of iGluRs include N-methyl-D-aspartic acid (NMDA: NR1-NR3) receptors (NMDARs), α-amino-3-hydroxy-5-methyl-4-isozazole propionic acid (AMPA) receptors (AMPARs) (GluA1-4), kainate receptors (GluKA1-2 and GluA5-7) and delta (δ1/δ2) receptors, which assemble into hetero- or homo-tetrameric structures (Dingledine et al., 1999). Overall, there are 18 different subunits, encoded by 6 distinct gene families based on sequence homology, shown in Figure **1.**

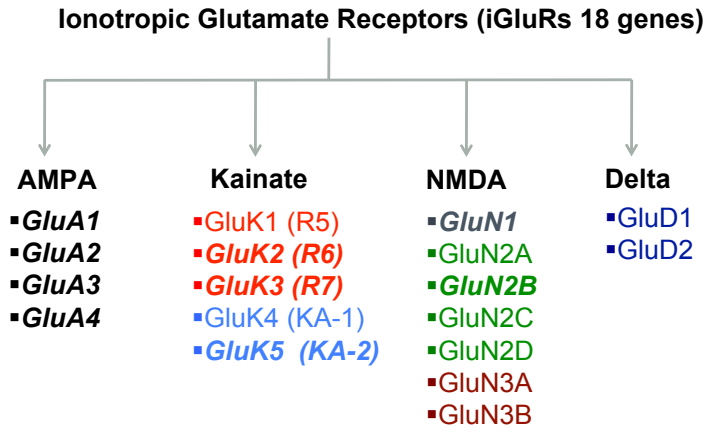


Figure **1**. Different sub-families of iGluRs .

(A) Overall, there are 18 different subunits, encoded by 6 distinct gene families based on sequence homology: 1 for AMPARs (*black*), 2 for Kainate (*red, blue*), and 3 for NMDARs (*grey, green, brown*). The ones in bold and italics have structurally resolved NTD, either in monomeric or dimeric forms.

Sequence similarity also suggests a similar architecture, comprising of an extracellular domain (ECD), a transmembrane domain (TMD or ion channel) and an intracellular carboxyl-terminal domain (CTD). The ECD of iGluRs consists, in turn, of a ligand-binding domain (LBD), and an N-terminal domain (NTD) (Traynelis et al., 2010) **(**Figure 2**A-C).**

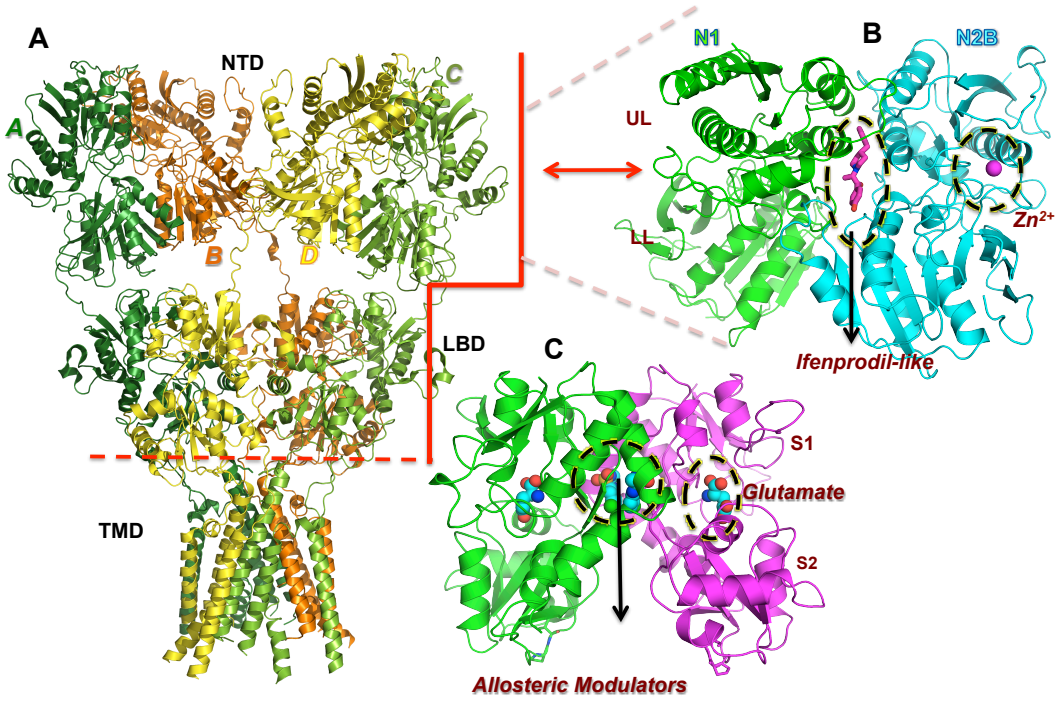


Figure 2. iGluR structure and known ligand binding sites.

A) Crystal structure of the intact iGluR belonging to the AMPAR subfamily. (B) Highlights the dimeric packing of the NTD of the NMDAR that is known to bind ligand. The clamshell-like structure divides the structure into upper lobe (UL) and lower lobe (LL). The binding site for two distinct allosteric regulators: polyamines like Ifenprodil that bind at the heterodimer interface (Karakas et al., 2011) and Zn2+ that binds one of the subunits (GluN2B (Karakas et al., 2009)) are shown here in circled regions. (C) Shows the dimeric packing of an AMPAR LBD. The LBD also has a bilobal clamshell-like architecture, dividing the LBD into S1 and S2 region with ligand-binding site at the cleft (encircled region, showing glutamate binding). Also, allosteric modulators of the LBD have been shown to bind at the dimer interface (Pohlsgaard et al., 2011).

So far, analyses of iGluRs have been concentrated on the isolated ECDs. Extensive experimental (Ahmed et al., 2007; Cheng et al., 2002; McFeeters and Oswald, 2002) and computational studies (Arinaminpathy et al., 2002; Lau and Roux, 2007; Lau and Roux, 2011; Pang et al., 2005) have focused on agonist/ligand binding to the LBD, driven by the availability of a host of structural data in the presence of various ligands (Pohlsgaard et al., 2011) (**Figure 3A**). A consensus from these studies is the clamshell-like closure of the bilobate structure of

LBD monomer (S1-S2) upon glutamate (or agonist) binding (**Figure 2C**). This motion allosterically triggers the opening of the TMD pore (Erreger et al., 2004; Mayer, 2006). The relaxation of the 'stressed' ligand-bound state is then achieved by receptor deactivation; through ligand unbinding or a ligand-bound but de-sensitized state (Armstrong et al., 2006; Hansen et al., 2007; Horning and Mayer, 2004; Sun et al., 2002). Allosteric modulators can alter desensitization characteristics by binding at the LBD dimer interface (**Figure 2C**). However, the exact mechanism of pore opening coupled to LBD lobe closure and consequent desensitization remains to be established.

While the structure and dynamics of the LBD have been fairly well studied, those of the NTD are less understood, primarily due to availability of less structural information. The main function of the NTD has so far been thought to be associated with mediation of synaptic protein-protein interactions and receptor assembly (Ayalon et al., 2005; Clayton et al., 2009; Greger et al., 2007; Hansen et al., 2010). However, the advent of new crystal structures for iGluR NTDs (Farina et al., 2011; Jin et al., 2009; Karakas et al., 2009; Karakas et al., 2011; Kumar and Mayer, 2010; Kumar et al., 2009; Rossmann et al., 2011; Yao et al., 2011), and the establishment of the allosteric behavior of NMDAR NTDs (Gielen et al., 2008; Gielen et al., 2009; Hansen et al., 2010; Mony et al., 2009) (**Figure 3B** shows current distribution of NTD structures) have now propelled studies towards understanding the dynamic landscape of NTDs and their potential role in channel gating.

Like, the LBD, the NTDs form stable dimers and feature the clamshell-like bilobate fold belonging to the periplasmic binding protein (PBP)-like family (O'Hara et al., 1993; Quiocho and Ledvina, 1996). To date, ligand-binding to this distal domain in iGluRs has been associated exclusively with the NMDA-type receptors (NMDARs): Zn^{2+} binding to GluN2 subunit NTDs

results in a down regulation of channel activity, presumably via closure of the NTD clamshell (Hansen et al., 2010; Karakas et al., 2009) (**Figure 2B** shows the Zn^{2+} binding site). Ifenprodil-like compounds were previously hypothesized to bind at NMDAR GluN2B subunit cleft based on docking and mutational analyses (Mony et al., 2009; Perin-Dureau et al., 2002). A more recent crystal structure of the heterodimer of GluN1 and GluN2B (Karakas et al., 2011), revealed that the binding site is, in fact, located at the interface of the two NTD monomers (**Figure 2B**). This recent structural data also underscores a mechanistic basis for this allosteric signalling in NMDARs where the lower NTD lobes are separated and thus free to move in response to ligand binding (Farina et al., 2011; Karakas et al., 2011). This arrangement also mimics the ligand binding cores of mGluRs, but is different from nonNMDA iGluRs. AMPAR and KR NTD dimers are usually 'zipped-up' across both lobes providing an extensive assembly interface (Kumar et al., 2009; Clayton et al., 2009, Jin et al., 2009; Rossmann et al., 2011), which likely restricts lobe motions. Accordingly, AMPA- and NMDA-R NTDs would have evolved different functions, with the AMPAR NTD mainly directing subunit assembly (Ayalon and Stern-Bach, 2001; Hansen et al., 2010; Rossmann et al., 2011) but the NMDAR NTD additionally modulating ion channel function. However, more recent crystal structures of AMPARs have shed some light on the diversity of dimeric packing across AMPARs and have enabled us to pursue a more comprehensive analysis of this elusive domain (described in Chapters 3-6).

The studies described above, are all for the isolated LBD or NTD. The determination of the crystal structure of the intact AMPAR by the Gouaux lab (Sobolevsky et al., 2009) revealed, for the first time, the modular architecture and atomic structure of the intact tetrameric receptor (except the CTD) in the closed form (of the TMD) (**Figure 2A**). The receptor has an overall two-fold axis of symmetry, with the NTD and LBD organized as two dimers while the TMD exhibits

four-fold symmetry. Also, domain swapping between the NTD and LBD leads to pairing of different subunits in the NTD (*AB, CD*) and the LBD (*AD, BC*) (**Figure 2A**). This hallmark structure offers a unique platform to analyze the collective dynamics of all three domains crystallographically resolved (NTD, LBD and TMD).

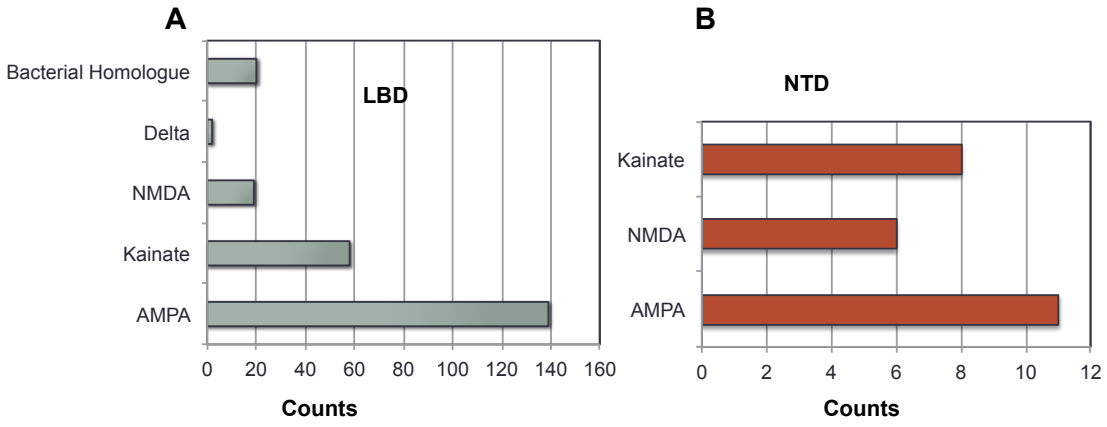


Figure 3. Distribution of available structures of iGluR NTD and LBD

A) Distribution of currently available crystal structures of eukaryotic and bacterial homologues of the LBD (B) Distribution of currently available crystal structures of the NTD.

1.2 OUTLINE OF THE DISSERTATION

The work in this dissertation is mainly focused on the study of structure and dynamics of iGluRs and allosteric potential in AMPARs. Our work is composed of a multi-scale approach, using both coarse-grained as well as all-atom models, to investigate the structural dynamics and ligand-binding propensities of iGluR NTDs (Chapters 3 and Chapter 5, respectively). We also unite the knowledge gained from structural dynamics with information from graph-theory approaches for identifying residues that play a key role in signal transmission. Focusing on the only crystallized member of the intact iGluR (AMPA - GluA2), we identify a previously uncharacterized

mechanism of possible NTD-mediated allosteric control (Chapter 4). Structural dynamics is closely related with evolutionary dynamics as obtained from multiple sequence alignments of homologous sequences. To this end, we developed an extension of existing software (*ProDy - evol*) to enable fast and efficient analysis of MSAs and evaluation of conservation and coevolution properties. Using this software, and other novel algorithms for coevolution analysis, we performed a hierarchical analysis of the iGluR NTD MSAs (Chapter 6). Our studies on iGluR, dynamics and allostery was conducted in collaboration with Prof. Ingo Greger and Prof. Indira Shrivastava and the work on development of *evol* and exploring the ligand binding of iGluR was done in collaboration with Dr. Ahmet Bakan.

The publications associated with the above studies are described below.

1. **Dutta A***, Bakan A* and Bahar I. Evol: Comparative Analysis of Protein Sequence Co-evolution and Structural Dynamics. (*under preparation*) **equal contribution*

2. **Dutta A** and Bahar I. A series of residues acting as sensors and effectors regulate allosteric communication in the intact AMPA receptor. (*under preparation*)

3. **Dutta A***, Shrivastava I.H*, Sukumaran M, Greger I.H and Bahar I. (2012). Comparative dynamics of NMDA- and AMPA-glutamate receptor N-terminal domains. Structure. 20: 1838-49 **equal contribution*

4. Eyal E, **Dutta A** and Bahar I. (2011). Cooperative dynamics of proteins unraveled by network models. WIREs Computational Molecular Science. 1: 426-39.

5. Sukumaran M, Rossmann M, Shrivastava I, **Dutta A**, Bahar I and Gregor I.H. (2011). Dynamics and allosteric potential of the AMPA receptor N-terminal domain. EMBO 30: 972-82.

6. **Dutta A** and Bahar I. (2010). Metal-binding sites are designed to achieve optimal mechanical and signaling properties. Structure. 18: 1140-8.

2.0 ELASTIC NETWORK MODELS: THEORY AND APPLICATION

Proteins possess the ability to undergo a distribution of collective changes in conformation, or modes of motions, at their equilibrium or native state, which accommodate, if not facilitate, their function (Bahar et al., 2007). Coarse-grained NMA methods have found widespread use in characterizing these collective motions. In particular, ENMs have found broad utility in conjunction with NMA, after the work of Tirion (Tirion, 1996), Bahar and coworkers (Bahar et al., 1997; Haliloglu et al., 1997) and Hinsen (Hinsen, 1998; Hinsen et al., 2000) that showed that cooperative movements that underlie many activities can be captured by network models (Chapter 1). The three major reasons behind the use of the ENMs are their simplicity, the robustness of the predicted modes of motions in the low frequency regime, also called the softest modes, and the functional significance of these modes indicated by numerous applications (Bahar et al., 2010a; Cui and Bahar, 2006).

In this chapter we outline the methodological details of ENMs, and other methods that are derived from ENM based characterization of protein 3D-structure. Its application towards understanding the specific features and communication propensities of metal-binding sites in enzymes are also presented here. The study provides us with insights into simple design principles. Functional residues/sites are implicated in some major way (e.g. hinge-bending, redistribution of salt bridges, conformational switches) in the softest motions, which are readily triggered by external perturbations (e.g. ligand binding).

2.1 NORMAL MODE ANALYSIS

Classical normal mode analysis (NMA) originated from spectroscopic analyses of molecules, where the absorption in the vibrational spectra is given by normal modes. Its first application to proteins dates back to early 80s (Brooks and Karplus, 1983; Go et al., 1983; Levitt et al., 1985). Renewed interest in the last decade stems from the fact that the low frequency modes can be robustly and efficiently derived using simplified models such as ENMs, and despite such simplifications the global modes accessible to biomolecular systems are robustly determined and they usually relate to functional motions (Bahar and Rader, 2005; Cui and Bahar, 2006).

The potential energy $V(\boldsymbol{q})$ of a molecular system of n degrees of freedom (e.g., atomic coordinates) may be approximated to the second order of the series expansion around the equilibrium state $\boldsymbol{q}^0 = [q_1^0 \ q_2^0 \ q_3^0 \ \ldots . \ q_n^0]^T$ as:

$$V(\boldsymbol{q}) \ = \ \frac{1}{2}\sum_{i,j}\left(q_i - q_i^0\right)\left[\frac{\partial^2 V}{\partial q_i \, q_j}\right]^0\left(q_j - q_j^0\right) \ = \ \frac{1}{2}\Delta\boldsymbol{q}^T \boldsymbol{K}\,\Delta\boldsymbol{q} \tag{1}$$

where $\Delta\boldsymbol{q}$ is the $3n$-dimensional vector of the instantaneous fluctuations, the superscript $\boldsymbol{T}$ designates its transpose, and $\boldsymbol{K}$ is the positive semi-definite matrix of second derivatives of the potential with respect to atomic coordinates. The ij^{th} element of $\boldsymbol{K}$ is $K_{ij} = [\partial^2 V/\partial q_i \, \partial q_j]^0$. If we treat structural points as harmonic oscillators in the absence of other effects, the protein obeys the equation of motion,

$$\boldsymbol{M}\frac{d^2\Delta\boldsymbol{q}}{dt^2} + \boldsymbol{K}\Delta\boldsymbol{q} \ = \ 0 \tag{2}$$

where $\boldsymbol{M}$ is a diagonal matrix composed of n super-elements $m_i\,\mathbf{I_3}$ along the diagonal, where m_i is the mass of the i^{th} atom and $\mathbf{I_3}$ is the identity matrix of order 3. The general solution to Eq. 2 is a $3n$-dimensional vector of the form $\Delta\boldsymbol{q}(t) = \boldsymbol{a}e^{i\omega t}$, which, upon substitution into Eq. 2, leads to

$$\left(-\omega^2 \boldsymbol{M} + \boldsymbol{K}\right)\boldsymbol{a} = 0 \tag{3}$$

Pre-multiplication of Eq. 3 by $\boldsymbol{M}^{-1/2}$ yields $\omega^2 \boldsymbol{M}^{1/2}\boldsymbol{a} = \boldsymbol{M}^{-1/2}\boldsymbol{K}\boldsymbol{a} = \boldsymbol{M}^{-1/2}\boldsymbol{K}[\boldsymbol{M}^{-1/2}\boldsymbol{M}^{1/2}]\boldsymbol{a}$, which, upon the change of variables $\boldsymbol{u} = \boldsymbol{M}^{1/2}\boldsymbol{a}$, $\lambda = \omega^2$, and $\boldsymbol{H} = \boldsymbol{M}^{-1/2}\boldsymbol{K}\boldsymbol{M}^{-1/2}$ leads to the eigenvalue equation

$$\lambda\ \boldsymbol{u} = \boldsymbol{H}\boldsymbol{u} \tag{4}$$

NMA is the solution of this equation to obtain the $3n$-6 nonzero eigenvectors $\boldsymbol{u}^{(k)}$ of the Hessian $\boldsymbol{H}$, along with the corresponding eigenvalues, λ_k. The eigenvectors are $3n$-dimensional vectors, the elements of which are organized in 3-dimensional vectors that represent the displacements $\boldsymbol{u}_1^{(k)}$, $\boldsymbol{u}_2^{(k)}$, ... $\boldsymbol{u}_n^{(k)}$ of the n atoms away from their equilibrium positions as the structure reconfigures along a given mode (e.g. mode k); and the eigenvalue λ_k is the corresponding squared frequency. λ_k scales with the curvature of the energy landscape along mode k. Thus, the lower frequency modes have a smaller curvature/stiffness and they undergo larger excursions from the energy minimum for a given energy increase, hence their 'soft modes' attribute.

2.1.1 ENM: Gaussian Network Model (GNM)

The use of ENMs to represent native proteins has two computational advantages: (i) there is no need for energy minimization; the Protein Data Bank (PDB) (Berman et al., 2000) structure is assumed to represent a global energy minimum, (ii) a coarse-grained model is adopted for structure and energetics: typically, each node represents a residue, and the pairs of nodes within a certain cut-off distance (R_c) are connected by springs of uniform force constant γ, which significantly reduce the size and complexity of $\boldsymbol{H}$. Two ENMs, the GNM (Bahar et al., 1997; Haliloglu et al., 1997) and the anisotropic network model (ANM) (Atilgan et al., 2001; Eyal et al., 2006) have found wide use in the last decade.

GNM. Following the original statistical thermodynamics theory of random polymer networks (Flory, 1976), the network nodes are assumed to undergo Gaussian fluctuations $\Delta \boldsymbol{R}_i$ ($1 \leq i \leq N$) about their mean positions. Likewise, the distance vectors between nodes undergo Gaussianly distributed fluctuations $\Delta \boldsymbol{R}_{ij} = \boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^0 = \Delta \boldsymbol{R}_j - \Delta \boldsymbol{R}_i$ (Figure 4**).** The coordinates of α-carbons are used to define the spatial position of the nodes. The GNM potential is given in terms of the fluctuations $\Delta \boldsymbol{R}_i = \boldsymbol{R}_i - \boldsymbol{R}_i^0$ in the position vectors of the nodes as[3]

$$V_{GNM} = \frac{\gamma}{2}\left[\sum_i \sum_{j,j>i} (-\Gamma_{ij})\left[\Delta \boldsymbol{R}_j - \Delta \boldsymbol{R}_i\right]^2\right] = \frac{\gamma}{2}\left[\Delta \boldsymbol{X}^T \boldsymbol{\Gamma} \Delta \boldsymbol{X} + \Delta \boldsymbol{Y}^T \boldsymbol{\Gamma} \Delta \boldsymbol{Y} + \Delta \boldsymbol{Z}^T \boldsymbol{\Gamma} \Delta \boldsymbol{Z}\right] \qquad (5)$$

Here $\boldsymbol{\Gamma}$ is the Kirchhoff (or connectivity) matrix, the off-diagonal elements of which are defined as $\Gamma_{ij} = -1$ if $|\boldsymbol{R}_{ij}| \leq R_c$ and zero otherwise; and the i^{th} diagonal elements is evaluated from the summation $\Gamma_{ii} = -\Sigma_j \Gamma_{ij}$ over all off-diagonal elements in the i^{th} row (or column); $\Delta \boldsymbol{X}$, $\Delta \boldsymbol{Y}$ and $\Delta \boldsymbol{Z}$ are the N-dimensional vectors of the respective X-, Y-, and Z-components of the fluctuation vectors $\Delta \boldsymbol{R}_1$, $\Delta \boldsymbol{R}_2$, .., and $\Delta \boldsymbol{R}_N$ corresponding to the N residues.

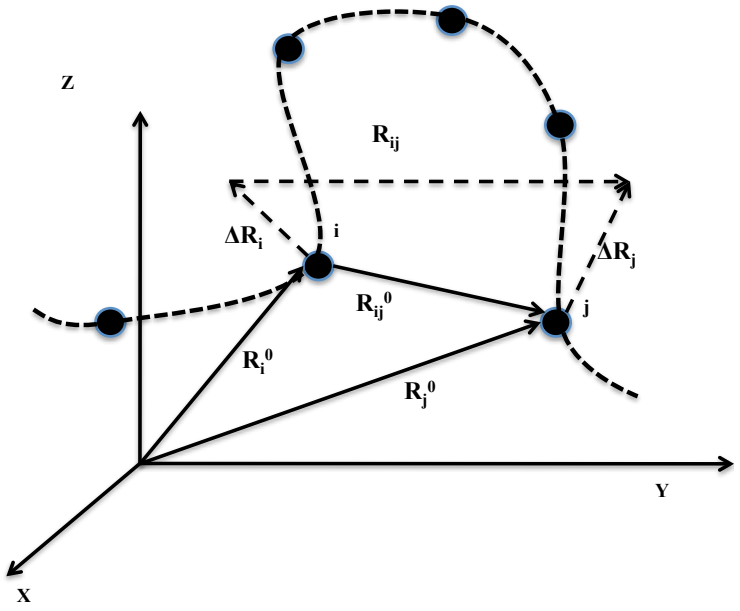


Figure 4. Schematic representation of equilibrium fluctuations

A portion of the protein backbone is displayed by the dotted curve. Filled dots refer to interaction sites (e.g., C^{a}-atoms) that are adopted as the network nodes. R_i^0 and R_j^0 designate the equilibrium positions of residues *i* and *j*; R_i and R_j are their instantaneous position vectors. The original and instantaneous separations are indicated by the solid line and dashed line, respectively. The fluctuations in the position vectors are given by ΔR_j and ΔR_i. R_{ij}^0 and R_{ij} designate the equilibrium and instantaneous distance vectors between residues *i* and *j*. The change in the inter-residue distance vector is related to the fluctuations in residue positions as $\Delta R_{ij} = R_{ij} - R_{ij}^0 = \Delta R_j - \Delta R_i$, and may be expressed as a weighted sum over the contributions $D R_{ij}^{(k)}$ of individual modes.

The cross-correlations between the fluctuations of residues *i* and *j* are found from the statistical mechanical average (Bahar et al., 1997)

$$\left\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_j \right\rangle = \int \left(\Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_j \right) exp\left(-V / k_B T \right) d\{\Delta \boldsymbol{R}\} \Big/ \int exp\left(-V / k_B T \right) d\{\Delta \boldsymbol{R}\} = \frac{3 k_B T}{\gamma} \left[\boldsymbol{\Gamma}^{-1} \right]_{ij} \quad (6)$$

where $[\Gamma^{-1}]_{ij}$ is the ij^{th} element of $\boldsymbol{\Gamma}^{-1}$, k_B is the Boltzmann constant and T is the absolute temperature. Note the determinant of $\boldsymbol{\Gamma}$ is 0, i.e., $\boldsymbol{\Gamma}$ cannot be inverted. Instead, its pseudoinverse is calculated using the N-1 nonzero eigenvalues σ_k and eigenvectors $\boldsymbol{v}^{(k)}$ of $\boldsymbol{\Gamma}$. In compact notation,

$$\boldsymbol{C}^{(N)} = = \frac{3k_BT}{\gamma}\boldsymbol{\Gamma}^{-1} = \frac{3k_BT}{\gamma}\sum_{k=1}^{N-1}\left[\sigma_k{}^{-1}\boldsymbol{v}^{(k)}\boldsymbol{v}^{(k)\mathrm{T}}\right] \tag{7}$$

Here $\boldsymbol{C}^{(N)}$ is the *NxN* covariance matrix; its ij^{th} element is $C_{ij}^{(N)} = <\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j>$, and the i^{th} diagonal element $C_{ii}^{(N)}$ is simply the mean-square fluctuations $<(\Delta\boldsymbol{R}_i)^2>$ of residue *i*. The contribution of any subset of modes to the cross-correlations or mean-square fluctuations may be evaluated by performing the summation in Eq 7 over this particular subset. Note that the soft modes (smallest eigenvalues) make the largest contribution to the covariance. The eigenvectors are normalized such that the plot of $[\boldsymbol{v}^{(k)}\,\boldsymbol{v}^{(k)\mathrm{T}}]_{ii}$ as a function of residue *i* represents the probability distribution of residue fluctuations in mode *k*, also called the k^{th} mode profile. The average "mobility profile" from contribution of *m* modes is given by:

$$\left\langle M_i \right\rangle|_m = \frac{\sum_{k=1}^{m}\sigma_k^{-1}[v^{(k)}]_i^2}{\sum_{k=1}^{m}\sigma_k^{-1}} \tag{8}$$

where the reciprocal σ_k^{-1} serves as the statistical weight of mode *k*. The above equation yields the contribution of the first *m* modes to the overall dynamics. This can also be used to calculate the contribution of any subset, e.g., the last 20 modes or the fastest modes.

Experimental properties like X-ray crystallographic B factors can be predicted using

$$\mathrm{B_i} = 8\pi^2\left(\frac{k_BT}{\gamma}\left[\boldsymbol{\Gamma}^{-1}\right]_{ii}\right) \tag{9}$$

This expression provides a good way to measure the level of agreement between GNM predictions and experimentally observed fluctuations of residues.

2.1.2 ENM: Anisotropic Network Model (ANM)

The ANM potential, V_{ANM}, is similar in form to V_{GNM}, except for the replacement of the scalar product $[\Delta \boldsymbol{R}_j - \Delta \boldsymbol{R}_i]^2 \equiv [(\boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^0) \bullet (\boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^0)]$ in Eq. 5 by $[|R_{ij}| - |R_{ij}^0|]^2$ where $|R_{ij}|$ designates the magnitude of R_{ij}. Thus, V_{GNM} penalizes the change in the orientation of $\Delta \boldsymbol{R}_{ij}$ even if the magnitude of the distance vectors is maintained; whereas V_{ANM} is exclusively based on distance changes (Atilgan et al., 2001). The 2nd derivative of V_{ANM} leads to expressions of the form

$$\partial^2 V_{ANM}/\partial X_i \partial Y_j \,|_{\boldsymbol{R}0} = -\gamma\, \Gamma_{ij} (X_j^0 - X_i^0)(Y_j^0 - Y_i^0)/[R_{ij}^c]^2 = -\gamma\, \Gamma_{ij} X_{ij}^0\, Y_{ij}^0 /[R_{ij}^0]^2 \qquad (10)$$

In terms of the components X_i^0, Y_i^0 and Z_i^0 of $\boldsymbol{R}_i^0$. This permits us to express the elements of ANM Hessian $\boldsymbol{H}$ in closed form, and easily evaluate the ANM eigenvectors $\boldsymbol{u}^{(k)}$ and eigenvalues l_k, $(1 \le k \le 3N\text{-}6)$ which provide information on the shape and frequencies of normal modes subject to V_{ANM}. $\boldsymbol{u}^{(k)}$ is a $3N$-dimensional vector, composed of 3-dimensional vectors $\boldsymbol{u}_1^{(k)}$, $\boldsymbol{u}_2^{(k)}$, ... $\boldsymbol{u}_N^{(k)}$ that describe the displacements of all N residues in mode k. The change in inter-residue distance induced by mode k is given in terms of the elements of $\boldsymbol{u}^{(k)}$ as

$$\Delta \boldsymbol{R}_{ij}^{(k)} = \boldsymbol{R}_{ij}^{(k)} - \boldsymbol{R}_{ij}^0 = (k_B T/\lambda_k)^{1/2} [\boldsymbol{u}_j^{(k)} - \boldsymbol{u}_i^{(k)}] \qquad (11)$$

Likewise, a $3N$x$3N$ covariance matrix, $\boldsymbol{C}^{(3N)}$, composed of blocks of size 3x3 associated with the three components of each fluctuation vector $\Delta \boldsymbol{R}_j$ replaces $\boldsymbol{C}^{(N)}$. $\boldsymbol{C}^{(3N)}$ scales with the inverse of the Hessian as

$$\boldsymbol{C}^{(3N)} = k_B T\, \boldsymbol{H}^{-1} = k_B T \sum_{k=1}^{3N-6} \left[\lambda_k^{-1} \boldsymbol{u}^{(k)} \boldsymbol{u}^{(k)\mathrm{T}} \right] \qquad (12)$$

To achieve a conformation displaced along one of the ANM modes, we use the following equation:

$$\mathbf{R}^{(k)} = \mathbf{R}^{(0)} \pm s\lambda_k^{-1/2}\, \mathbf{u}_k, \qquad (13)$$

where $\mathbf{R}^0$ is a 3 dimensional vector representing initial co-ordinates. Using different values of *s* we can generate an ensemble of conformations along mode *k*. *s* scales with k_BT/γ, where k_B is the Boltzmann constant, T is temperature and γ is the spring constant. γ can be approximated to reproduce experimental fluctuations from B factors with $B_i = (8\pi^2 k_B T \, tr[\mathbf{H}^{-1}]_{ii}) / 3\gamma$.

Overlap between predicted modes and experimentally observed structural changes: Consider two known structures A and B for the protein (family) of interest, represented each by the optimally superimposed 3*N*-d conformational vectors $\{\mathbf{R}_0\}_A$ and $\{\mathbf{R}_0\}_B$. Their structural difference (after optimal alignment) is $\Delta\{\mathbf{R}\}_{A\rightarrow B} = \{\mathbf{R}_0\}_B - \{\mathbf{R}_0\}_A$. To observe if A or B is pre-disposed to undergo this change, we evaluate the *overlap* or correlation cosine, $\boldsymbol{d} \cdot \mathbf{u}^{(k)}$, for a subset of soft modes $\mathbf{u}^{(k)}$ accessible to A (or for B). Here $\boldsymbol{d}$ is the unit vector along $\Delta\{\mathbf{R}\}_{A\rightarrow B}$. The cumulative overlap achieved by a subset of *m* modes is given by $CO(m) = [\Sigma_k (\boldsymbol{d} \cdot \mathbf{u}^{(k)})]^{1/2}$ where the summation is performed over $1 \le k \le m$. The complete set of eigenvectors form an orthonormal basis, i.e. $CO(m) = 1$ for $m = 3N-6$. Of interest is to see if $\Delta\{\mathbf{R}\}_{A\rightarrow B}$ is attainable via a small subset of soft modes, in which case the alternative structure simply represents a reconfiguration easily sampled in the ensemble of conformers intrinsically accessible to A.

2.1.3 Limitation

The GNM/ANM results depend on the overall 'fold' rather than specific interactions or detailed atomic coordinates. As such, they provide information on the global dynamics favored by the particular fold/architecture, rather than local changes in structure and interactions. The main utility of GNM/ANM is indeed the ability to efficiently explore global dynamics (collective motions in the low frequency regime), which have been demonstrated in multiple studies to be

insensitive to structural details or detailed force fields. Results on the high frequency modes have limited utility: they permit us to identify potentially conserved sites and folding nuclei, but provide no information on the mechanisms of local motions (Bahar et al., 1998; Haliloglu et al., 2005). It should also be noted that protein conformers used as input in, or generated as outputs from, ENM-NMA are not energy minimized with a molecular mechanics force field and hence may not be directly utilized in docking studies that require atomic level precision. Efforts are underway for designing hybrid methodologies that combine the information on global dynamics from ENMs and local motions from molecular dynamics simulations (Isin et al., 2008), to efficiently explore multiscale processes. Also sequence information is not included in ENMs and hence single mutations that result in loss of function are not captured by ENM analysis of the mutant, unless there are structural data for the mutant, which differs from that of the wild type protein. Clearly NMAs with ENMs also suffer from the same limitations (e.g. inadequate description of non-linear effects, solvent effect, etc) as atomic NMA, although coarse-graining of the structure and interactions allows for efficient sampling of domain rearrangements and cooperative movements near native state conditions that could be hardly accessible (due to local energy barriers) should a detailed, full atomic description of the structure and dynamics be implemented.

2.2 MARKOVIAN STOCHASTIC MODEL

Network models provide a useful tool for investigating the communication and signaling properties of complex systems. A widely used approach is to model signal transduction as a Markovian stochastic process. The Bahar laboratory recently explored the possibility of probing

the stochastics of signal propagation in proteins using a discrete time, discrete state Markov model (Chennubhotla and Bahar, 2007). The strengths of interactions (or weight of edges) in this network model (or the nonzero elements of **Γ**) could be assumed to be uniform (as in the GNM) as a first approximation. A slightly refined model, however, is to take account of residue specificity at a coarse-grained scale. To this aim, we define affinities $a_{ij} = N_{ij}/\sqrt{(N_i N_j)}$ as the weights of the edges, where N_{ij} is the total number of heavy atom contacts between residues *i* and *j* with the denominator correcting for bias due to size (**Figure 5**). The degree d_j of each node in is given by $\Sigma_i a_{ij}$. The Kirchhoff matrix corresponding to this model is then **Γ** = **D** - **A**, where **D** is the diagonal matrix of the degrees, and **A** is the matrix of affinities. The conditional probability of transmitting a signal from residue *j* to residue *i* in one time step is simply $m_{ij} = d_j^{-1} a_{ij}$ (where m_{ij} is the ij^{th} element of the Markov transition matrix **M**) (Chennubhotla and Bahar, 2007).

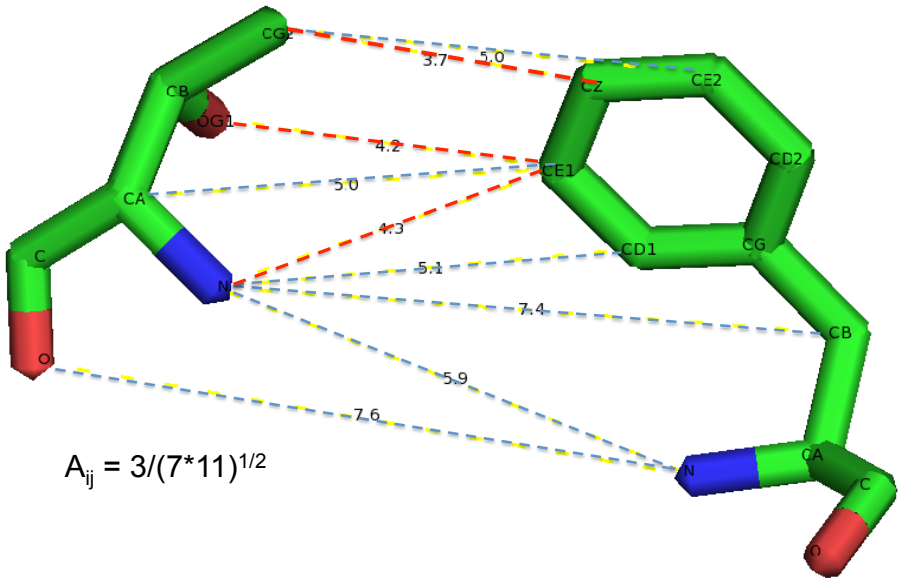


Figure 5. Schematic description of the evaluation of affinity matrix elements.

Two residues within interaction range are displayed, composed of $N_i = 7$ and $N_j = 11$ atoms. The affinity is evaluated based on atom-atom contacts that are closer than a cutoff separation, e.g. 4.0 Å. In the present case there is only one pair of atom within this interaction rage, such that $N_{ij} = 1$, and the affinity between this pair becomes $a_{ij} = N_{ij}/\sqrt{(N_i N_j)} = 1/(77)^{½}$

A metric of efficiency of signal transmission in network models is the hit time, which is the average number of steps it takes to transmit information from broadcaster *i*, to receiver *j* *(H(i,j))*. This is described by enumerating all possible ways to get from residue *i* to residue *j* weighted by the transition/conditional probability of signal flow. The recursive equation to evaluate the average hit time between nodes *i* and *j* is given by:

$$H(j,i) = \sum_{k=1}^{n}[1 + H(j,k)]m_{ki} = \sum_{k=1}^{n} m_{ki} + \sum_{k=1,k\neq j}^{n} H(j,k)m_{ki} = 1 + \sum_{k=1,k\neq j}^{n} H(j,k)m_{ki} \tag{14}$$

Substituting $\mathbf{M} = \mathbf{D}\mathbf{A}^{-1}$ and $\mathbf{\Gamma} = \mathbf{D} - \mathbf{A}$, *H(j,i)* can be expressed in terms of the elements of $\mathbf{\Gamma}^{-1}$ as(Chennubhotla and Bahar, 2007)

$$H(j,i) = \sum_{k=1}^{N}\left\{\left[\mathbf{\Gamma}^{-1}\right]_{ki} - \left[\mathbf{\Gamma}^{-1}\right]_{ji} - \left[\mathbf{\Gamma}^{-1}\right]_{kj} + \left[\mathbf{\Gamma}^{-1}\right]_{jj}\right\}d_k , \tag{15}$$

Or, using Eq. 6,

$$H(j,i) = \left(3k_BT/\gamma\right)^{-1}\sum_{k=1}^{N}\left[\left\langle \Delta R_k . \Delta R_i \right\rangle - \left\langle \Delta R_j . \Delta R_i \right\rangle - \left\langle \Delta R_k . \Delta R_j \right\rangle + \left\langle \Delta R_j . \Delta R_j \right\rangle\right] d_k \tag{16}$$

H(j,i) depends on the direction of signal transmission, i.e. $H(j,i) \neq H(i,j)$. Commute time, $\tau(i,j) \equiv H(j,i) + H(i,j)$, is another metric, which has no directionality.[14] $\tau(i,j)$ scales with the fluctuations in the distance vector $\Delta\boldsymbol{R}_{ij}$ as

$$\tau(i,j) \propto \left\{\left[\mathbf{\Gamma}^{-1}\right]_{ii} + \left[\mathbf{\Gamma}^{-1}\right]_{jj} - 2\left[\mathbf{\Gamma}^{-1}\right]_{ij}\right\} = \left\langle \Delta\boldsymbol{R}_{ij} \, . \, \Delta\boldsymbol{R}_{ij} \right\rangle \tag{17}$$

where the proportionality constant is $(3k_BT/\gamma)^{-1}\,\mathbf{\Sigma}_k\, d_k$. Eq. 17 is readily obtained using Eq. 15 twice, for *H(j,i)* and *H(i,j)*. Eqs. 14 and 15 permit us to bridge between statistical physical quantities such as mean-square fluctuations or cross-correlations with graph-theoretic concepts such as hitting or commute times. We will illustrate the utility of these concepts for analyzing

signal transduction behavior of proteins by way of application to metal binding proteins in Section 2.4.

2.3 NETWORK MODEL BASED PERTURBATION ANALYSIS

2.3.1 NMA of a subsystem coupled to an environment

Using above mentioned techniques, we capture the dynamics of an individual isolated system. However, to assess environmental effects, we use this method where the dynamics of a system (S) in the context of an environment (E) is evaluated by partitioning **H** into four submatrices (Ming and Wall, 2005; Zheng and Brooks, 2005)

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{SS} & \mathbf{H}_{SE} \\ \mathbf{H}_{ES} & \mathbf{H}_{EE} \end{pmatrix} \tag{18}$$

where $\mathbf{H}_{SS}$ refers to interactions within the system, $\mathbf{H}_{EE}$ to those within the environment, and $\mathbf{H}_{SE}$ (or $\mathbf{H}_{ES}$) to the coupling between S and E. The resulting effective Hessian of the system is in the presence of the environment

$$\mathbf{H}_{SS}^{eff} = \mathbf{H}_{SS} - \mathbf{H}_{SE}\mathbf{H}_{EE}^{-1}\mathbf{H}_{ES} \tag{19}$$

Decomposition of this effective hessian gives us the normal modes of the relevant subsystem encompassing the effects of the environment. Also, the eigenvectors obtained from system alone can be compared to those from subsystems by evaluating the correlation cosine [$\boldsymbol{u}^{(k)}$. $\boldsymbol{u}_{eff}^{(l)}$] between the eigenvectors $\boldsymbol{u}^{(k)}$ to $\boldsymbol{u}_{eff}^{(l)}$ (for k, l=1-40) corresponding to $\mathbf{H}_{SS}$ and $\mathbf{H}^{eff}_{SS}$, respectively.

2.3.2 Perturbation Response Scanning

This technique measures the displacement in a protein resulting from a perturbation at a single residue (Atilgan and Atilgan, 2009). The theory derives from Hooke's law, where f (*force*) = k (*force constant*)*Δr (*displacement*). Considering the ANM network model of C^{α} nodes connected by harmonic springs, the hessian (**H**) (see equation 10) gives the force constant for the system and consequently, $\mathbf{F} = \mathbf{H}\,\Delta\mathbf{R}$ or $\Delta\mathbf{R} = \mathbf{H}^{-1}\mathbf{F}$. We measure the displacement, $\Delta\mathbf{R}^{(i)}$ as a result of perturbation, modeled as force applied to residue *i*. This is called a scanning technique since at any one time we apply force on one residue $\mathbf{F}^{(i)} = [0,0,0, \ldots F_{xi}, F_{yi}, F_{zi}, \ldots 0,0,0]^{T}$ and repeat for all residues. For each residue *i* we apply unit force and measure the displacement of the other (N) residues in x, y, z direction making $\Delta\mathbf{R}^{(i)}$ a 3N-dimensional vector. The magnitude of response for residue *j* when unit force is applied at *i* is then calculated as a square displacement, such that $||\Delta\mathbf{R}_{j}^{(i)}||^2 = (\Delta r_{jx}^{(i)})^2 + (\Delta r_{jy}^{(i)})^2 + (\Delta r_{jy}^{(i)})^2$. This process is repeated *k* times so that the direction of forces are randomized *(+ve* or *–ve* directions), such that the final response is calculated by averaging over *k* instances and $\langle||\Delta\mathbf{R}_{j}^{(i)}||^2\rangle = (\Sigma_k(||\Delta\mathbf{R}_{j}^{(i)}||^2)_k)/k$.

This gives us the elements of the NxN PRS matrix where the ij^{th} element gives the displacement of residue *j* as a result of perturbation at residue *i* with unit forces averaged over *k* iterations. We further normalize the *ij*th element of the PRS matrix by dividing each row with the diagonal element. The PRS matrix obtained is an asymmetric matrix, where the rows give us the profile of "the effectiveness" of perturbation at residues and the columns of the matrix gives us the profile of "sensitivity" to perturbation (See Chapter 4, Section 4.2).

2.4 EXAMPLE CASE EXEMPLIFYING THE APPLICATION OF ABOVE METHODS FOR DISCERNING IMPORTANT PHYSICAL PROPERTIES

In a recent study of the relation between soft modes and catalytic activity for a series of enzymes, Yang and Bahar (2005) found that catalytic residues tend to be located near *key mechanical sites*. Here key mechanical sites refer to hinge sites or anchors that mediate the softest modes of motions, also called *global modes*, as predicted by the GNM (Bahar et al., 1997; Haliloglu et al., 1997). The probability of finding a catalytic residue among these key mechanical sites turns out to be 3–4 times higher than that from a random search. Such spatial proximity has been proposed to be a prerequisite for efficient coupling between chemical and mechanical activities (Yang and Bahar, 2005). A further study of signal propagation pathways using a Markovian stochastic model showed that allosteric structures are predisposed to instigate efficient communication mechanisms favored by their inherent network topology (Chennubhotla and Bahar, 2007). Following a similar premise, the following section, further exemplifies the importance of ENM based analyses (described in Sections 2.1 – 2.3) in understanding important physical characteristics of metal-binding sites.

2.4.1 Metal-Binding sites are designed to achieve optimal mechanical and signaling properties

Metal-binding proteins are associated with a variety of cellular functions. Some of them play roles in transport and cell signaling; others act as cofactors that are incorporated into enzymes to provide structural support while stabilizing functional conformations, or they directly participate in chemical reactions during enzyme catalysis (Kendrik et al., 1992; Tainer et al., 1992). A

number of web servers and databases (DBs) have been developed to annotate metal-binding proteins (Babor et al., 2008); and tools based on machine learning techniques such as support vector machines, neural networks and Bayesian classifiers have been developed for predicting metal-binding sites (Ebert and Altman, 2008; Lin et al., 2005; Lin et al., 2006; Passerini et al., 2007; Passerini et al., 2006) using the known structures of bound or '*holo'* proteins as training dataset. Edelman, Sobolev and coworkers (2008) developed such an algorithm for predicting binding sites for transition metals, i.e., those associated with "catalytic, co-catalytic or structural" roles (Babor et al., 2008). These metals are coordinated by three or more amino acids. Cysteine (C), histidine (H), glutamatic acid (E) and aspartic acid (D) are the most frequently observed metal-binding residues (Auld, 2001; Babor et al., 2008; Golovin et al., 2005), and are referred to as the "CHED" category of metal-binding residues.

In this study, we have analyzed the collective dynamics of a representative set of metal-binding proteins. The questions we asked were: do metal-binding sites possess structure-induced dynamic properties that enable their involvement/assistance in many activities? Are they distinguished by specific features or specific positions in the structure, which enable them to achieve their role in the activity of the proteins that bind them? Do they share some common design features (apart from those structurally known) that confer efficient cooperation and communication across the proteins?

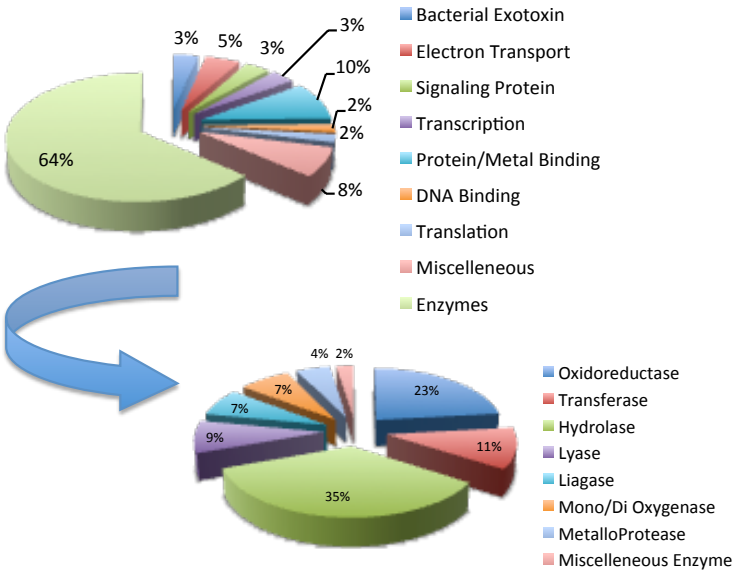


Figure 6. Functional distribution of *holo* proteins in the database of metal-binding proteins.

The dataset contains 145 *holo* proteins used by **(Babor et al., 2008)**. The majority of these structures are enzymes, and their distribution among different enzymatic classes is shown in the lower pie chart. Functional annotation was done using UniProt and PDB (Berman et al., 2000; Jain et al., 2009).

A representative dataset of metal-binding proteins previously used by (Babor et al., 2008) has been adopted here for a systematic analysis. **Figure 6** shows the broad range of functions represented by these proteins. 64% of them are enzymes with a diversity of biochemical activities, and the remaining 36% are almost equally distributed among other functions. We examined several properties of metal-binding residues, and compared them to those of the non metal-binding amino acids of the same type. First, we focused on the soft modes predicted by the GNM, to see the dynamic role assumed by metal-binding sites (apart from their chemical role of coordinating the ligand). Second, we examined their solvent accessibility. Third, using information-theoretic spectral methods, we mapped the signal flow pathways inherently accessible to these structures, and see if/how metal-binding sites are involved in establishing allosteric communication. The results are presented here for a set of 175 *holo* proteins where we

demonstrate that metal-binding residues occupy low mobility regions in the global modes. We illustrate that the stabilization of the bound metal results not only from local geometry and energetics, but from a global optimization of the intrinsic dynamics of the overall protein. These residues also tend to be buried in the structure, despite being polar or charged. Our study further shows that metal-binding sites serve as efficient signal transduction centers, suggesting that their particular location on the 3-dimensional structure has been evolutionarily optimized to achieve cooperative effects. The observed propensities provide guidelines for designing potential metal-binding sites in proteins, which are verified to be fulfilled by *de novo* metal-binding proteins.

2.4.2 Datasets

The analysis has been performed using 175 metal-binding proteins' structures deposited in the PDB (Berman et al., 2000). Sixty of these structures refer to metal-binding proteins that have been resolved in both *apo* and *holo* (with metal) forms (30 pairs). Dataset I in Appendix A lists the PDB codes, chain identifiers, lengths (number of resolved residues, *N*) of these structures, along with the identities of bound metals and metal-binding residues, and the root-mean-square deviations (RMSDs) between the two forms, both for the backbone and the metal-binding site. The RMSDs averaged over all pairs are 0.389 ± 0.351Å and 0.221 ± 0.332Å for the backbone and metal-binding sites, respectively (*Appendix A*), indicating that the proteins exhibit minor changes in structure upon metal binding. Datasets II and III include an additional 115 metal-binding proteins structurally resolved in *holo* form only (*Appendix A*). The complete set of 145 *holo* structures include all those compiled by Edelman et al (Babor et al., 2008), except for those whose ligand-binding sites could not be identified/verified using the *MetalloProtein Database* (Castagnetto et al., 2002), or those which have more than 90% sequence similarity with respect

to a member of the dataset. The bound metals include Zn (most frequent), Co, Ni, Fe, Mn and Cu.

2.4.3 *Apo* and *holo* forms exhibit similar global dynamics

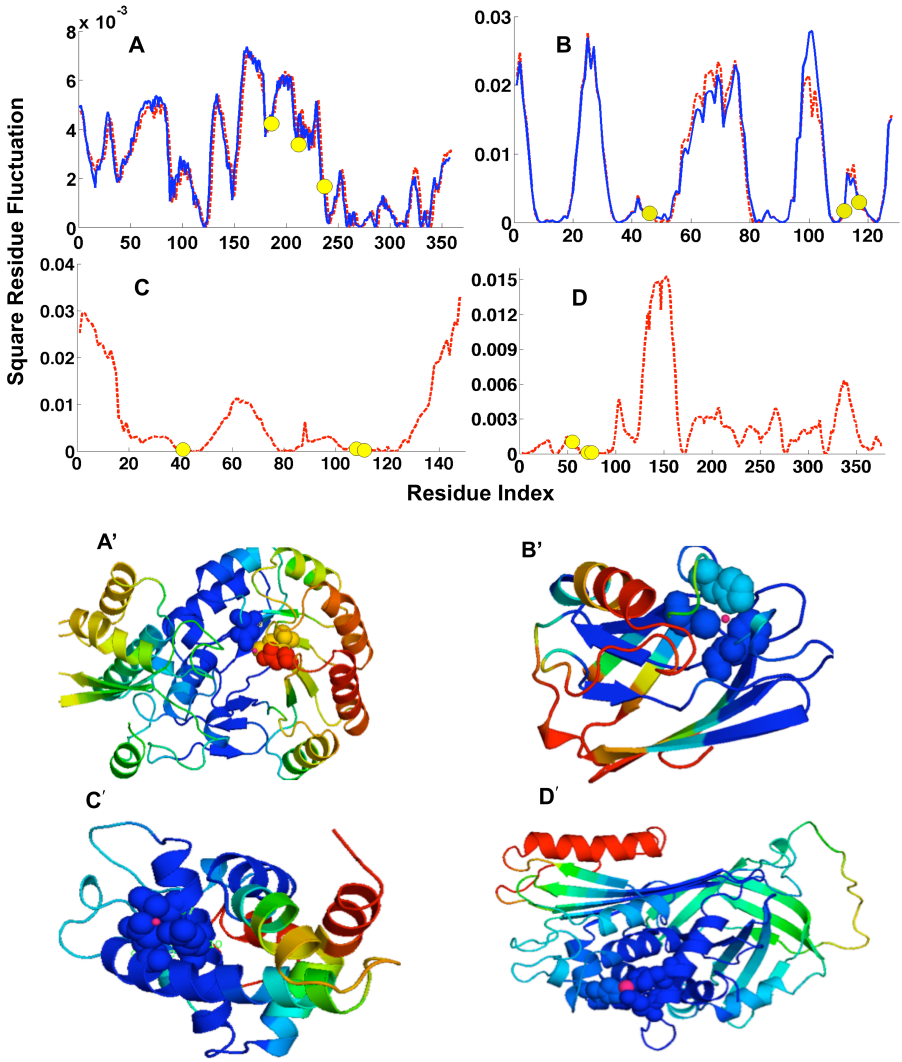


Figure 7. Global dynamics of metal-binding proteins illustrated for four cases.

A-D. Fluctuation profiles obtained by the GNM for four metal-binding proteins in *holo* form: A. 1MUC a muconate lactonizing enzyme with bound Mn^{2+}; B. 1VLX an electron transport protein with bound Co^{2+}; C. 1JFZ an RNase III endonuclease with bound Mn^{2+}; and D. 1HP7 an anti-trypsin binding Zn^{2+}. The curves represent the normalized distributions, or histograms, of square fluctuations, as a function of residue number, in the softest modes accessible to each structure. The yellow markers show the loci of metal-binding residues. These tend to occupy positions near

local or global minima. Panels **A** and **B** compare the profiles for the *holo* (red dashed curve) and *apo* (blue curve) forms and illlustrate that the two forms show minimal, if any, change in their global mode profile. **A', B', C', D'.** Ribbon diagrams of the four proteins in their *holo* forms, color-coded according to GNM softest mode profiles in panels **A-D**, from blue (most rigid) to red (most mobile). The metal-binding sites are shown in space-filling representation, and the metal ions in pink. Note that metal-binding sites are highly constrained in general (shown in blue), except for the structure in panel A/A'.

Figure 7A-D illustrates the global mobility profiles of a few metal-binding proteins in the softest modes of motions. Global mobility profiles refer to the normalized distributions of the square displacements of residues in the lowest frequency GNM mode. Mobile regions appear as peaks, whereas minima are regions with restricted movements that often pack functional residues in well-defined geometries. The panels **A'-D'** display the color-coded ribbon diagrams for the respective proteins. Metal-binding residues are indicated by filled circles (panels **A-D**) and displayed in space-filling representation (panels **A'-D'**).

Panels **A** and **B** of Figure **7** compare each the global mobility profiles obtained for the holo and *apo* forms of the same protein. The close superposition of the pairs of curves in each panel suggests that there is no observable difference between the global mode profiles of the metal-bound and -free forms. This trend is seen in practically all of the 30 pairs of structures resolved in the presence and absence of bound metal (Dataset I). The last two columns in the table containing Dataset I contains (i) the correlations between the global mobility profiles of the two forms, and (ii) the correlations between the X-ray crystallographic B-factors experimentally observed for the two forms. An average correlation of 0.936 is obtained between the global mobilities of the *apo* and *holo* forms; whereas their B-factors, which scale with the mean-square fluctuations (MSFs) or residues as $B_i = (8\pi^2/3) <(\Delta R_i)^2>$ exhibit an average correlation of 0.745. These results indicate there are some differences in the MSFs of residues in different forms, which may arise from minor structural differences between the two forms as well as different

packing geometry (intermolecular contacts) and other effects such as static disorder in the crystal structures. The global mobilities, on the other hand, are insensitive to small differences in structure, consistent with the well-established robustness of softest modes (Nicolay and Sanejouand, 2006; Tama and Brooks, 2006).

These results confirm that the global dynamics of a given protein is a collective property of its overall architecture, and ligand/metal binding has minimal, if any, effect on its intrinsically accessible soft motions, in accord with previous experimental and computational observations made for liganded and unliganded forms of enzymes (Bakan and Bahar, 2009; Eisenmesser et al., 2005; Lange et al., 2008; Tobi and Bahar, 2005; Yang and Bahar, 2005). In view of the insensitivity of the global modes to the presence/absence of bound metal, we focus on the dynamics of 145 *holo* proteins listed in Datasets II and III.

2.4.4 Metal-Binding sites have restricted fluctuations

Next, we examined whether metal-binding sites occupy positions coinciding with, or close to, key mechanical sites in the 3-dimensional structure of the proteins. Key mechanical sites serve as hinges/anchors in the global modes, and as such they appear as minima in the global mode profiles. Panels **A-D** in **Figure 7** indicate a tendency of metal-binding sites (indicated by yellow circles) to be located near minima (local or global), although this trend is not that distinctive. Likewise, the color-coded diagrams in panels **A′-D′** also indicate relatively low mobilities (blue regions) at metal-binding residues (shown in space-filling representation), although departures from this behavior are also observable (e.g., panel A').

Toward a more critical assessment of the relationship, if any, between metal-binding sites and key mechanical sites, we performed a comparative analysis of the mobilities of three groups

of residues: all residues, metal-binding CHED (Cys, His, Glu and Asp) residues and all other CHED residues in the Datasets I-III.

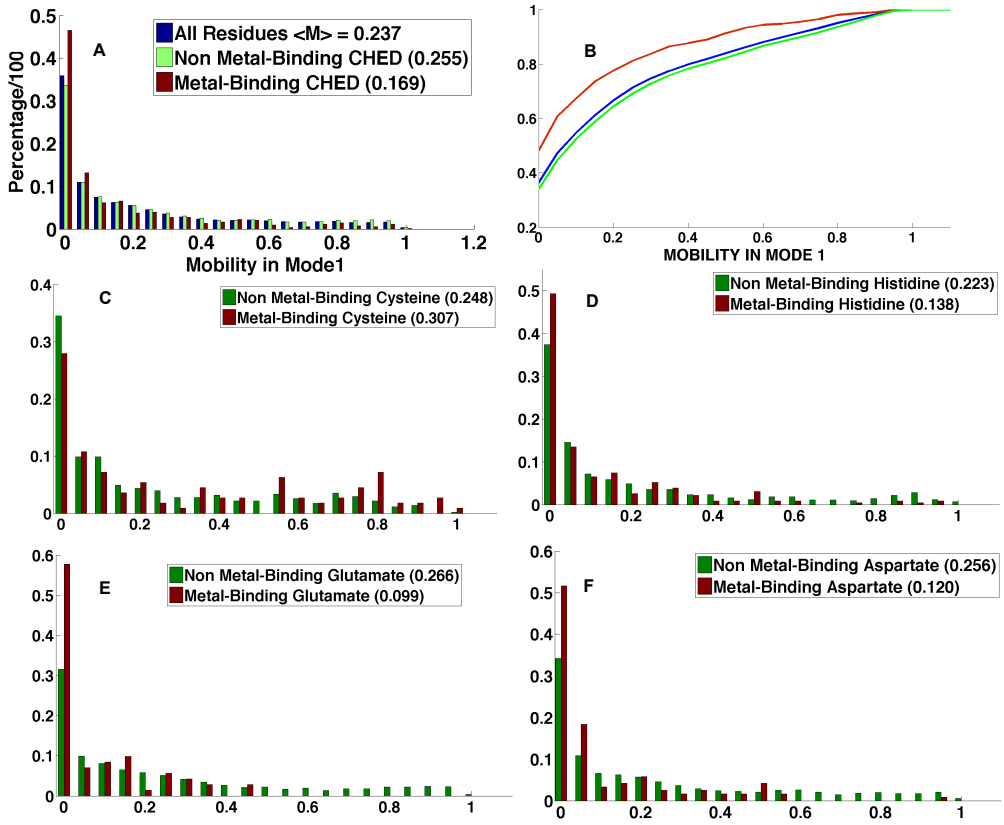


Figure 8. Comparison of the global mobilities of different types of residues.

A. Histograms of mobilities for three different groups of residues: metal-binding CHED, all CHED (green) and all (blue) residues. Mobilities are normalized in the range [0, 1], and results are shown for 21 bins at an interval of 0.05; the first bin refers to the count of residues having mobilities in the range from 0.00 to 0.05 and so on, expressed as probabilistic occurrence on the ordinate. **B.** Cumulative distributions of mobilities for the same three groups of residues. C-F. Same as in panel **A**, for specific amino acids at metal-binding and other (green) locations. Metal-binding '*HED*' residues exhibit mobilities significantly lower than their non-metal-binding counterparts, while cysteines ('*C*') show the opposite trend. The numbers in parenthesis in the insets show average mobilities.

The results are presented in **Figure 8**. Panel **A** displays the histograms of mobilities for the three subsets. For comparative purposes, the mobilities were normalized in the range [0, 1] for each protein. Metal-binding CHED residues, indicated by the red bars, exhibit a clear bias towards lower mobilities compared to the residues in the other two groups. This is despite the

fact that CHED residues are charged or polar, and tend to positioned on the surface of the protein, thus enjoying higher mobility as compared to other residues. The mean values and variances corresponding to the three distributions are 0.24 ± 0.27 for all residues and 0.26 ± 0.28 for non-metal-binding CHED, and 0.17 ± 0.22 for metal-binding CHED (Table S4). The variance is high because of the long-tail of the distributions. Panel **B** presents the same results as cumulative distributions. Almost 65% of metal-binding CHED residues have mobilities lower than 0.1.

Table 1. Fractional Mobility Statistics for Different Types of Residues.

TYPE OF RESIDUE	SIZE	OBSERVED MOBILITY		EXPECTED MOBILITY	
		MEAN	COVARIANCE	MEAN	COVARIANCE
All residues	44,291	0.24	0.27	-	-
Non Metal-Binding CHED residues	7170	0.26	0.28	0.25	0.28
Metal-Binding CHED Residues	518	0.17	0.22	-	-
Metal-Binding HISTIDINE	227	0.14	0.19	-	-
Metal-Binding ASPARTIC ACID	119	0.12	0.17	-	-
Metal-Binding GLUTAMIC ACID	69	0.10	0.12	-	-
Metal-Binding CYSTEINE	103	**0.31**	**0.31**	-	-
Non Metal-Binding HISTIDINE	1044	0.22	0.28	0.22	0.28
Non Metal-Binding ASPARTIC ACID	2656	0.26	0.28	0.24	0.28
Non Metal-Binding GLUTAMIC ACID	2966	0.27	0.29	0.27	0.29
Non Metal-Binding CYSTEINE	504	**0.25**	**0.27**	0.22	0.26

In **Figure 8** panels **C-F**, we take a closer look at each type of CHED residues and compare the mobilities of the metal-binding and non-metal-binding subsets. Among these four types of amino acids, it is interesting to note that cysteines exhibit a fundamentally different behavior: while metal-binding His, Glu and Asp possess a significantly lower mobility (in the *holo* forms) compared to their non-metal-binding counterparts, metal-binding cysteines enjoy a higher mobility than those not involved in metal-binding. **Table 1** summarizes the mean mobilities and their standard deviation data for group of each amino acid. Essentially, the metal-binding HED residues display mean mobilities around 0.12 (smaller than the average mobility of all residues by a factor of 2), while non-metal binding HED residues are almost indistinguishable from all other residues. Cysteines exhibit the opposite pattern: metal-binding cysteines are even more mobile than an average non-metal-binding residue.

The statistical significance of the results has been tested using three methods: (i) upon examining the distribution of the mobilities of an ensemble of randomly selected residues (of the same size as the metal-binding CHED residues), which showed that the mean and covariance (0.24 ± 0.27) are comparable to those of the entire ensemble, repeated and confirmed for five independent runs; (ii) performing the same type of analyses for each of the CHED residues, which indicated the distinctive properties of metal-binding CHED (**Table 1**), and (ii) doing a KS test at 5% level of significance to verify that metal-binding CHED residues have a statistically significant difference in their mobility distribution compared to their non-metal-binding counterparts (**Table 2**), except for the subset of Cys residues.

Table 2. Sample KS test for the statistical significance of differences observed between the metal-binding and non metal-binding residues' mobility distributions.

Distribution1	Distribution2	Test result -H	p-value	Null Hypothesis	Same Distribution?
All Residues	Metal-Binding CHED	1	$2.86*10^{-9}$	Reject	No
All Residues	Non Metal Binding CHED	0	0.061	Accept	Yes
Metal-Binding CHED	Non Metal Binding CHED	1	$4.79*10^{-12}$	Reject	No
Histidine- metal-binding	Histidine- not metal-binding	1	$9.57*10^{-4}$	Reject	No
Cysteine- metal-binding	Cysteine- not metal-binding	0	0.1709	Reject	No
Glutamate- metal-binding	Glutamate- not metal-binding	1	$1.17*10^{-5}$	Reject	No
Aspartate- metal-binding	Aspartate- not metal-binding	1	$6.17*10^{-7}$	Reject	No

2.4.5 Metal-binding sites show decreased solvent accessibility

A comparative analysis of solvent accessibility (Eyal et al., 2004; McConkey et al., 2002) properties sheds light on the distinctive properties of the two subgroups of CHED residues: Metal-binding CHED residues have much smaller solvent accessible surface areas (SASAs) compared to non-metal-binding CHED, as may be viewed from panels **A** and **B** in **Figure 9**. Panels **D-F** shows that there is a drastic difference in the solvent exposure of histidines, glutamates and aspartates belonging to the metal-binding and non-metal-binding subgroups: the former subgroup exhibits a distinctive preference for more buried positions. Notably, ~ 80% of metal-binding glutamates have normalized SASAs smaller than 0.05, as compared to 10% of the non metal-binding glutamates. Cysteines (panel **C**), on the other hand, do not appear to

differentiate between the two subgroups: they tend to occupy buried positions irrespective of metal binding. About 60% of both metal-binding and non metal-binding cysteines have normalized SASAs lower than 0.05.

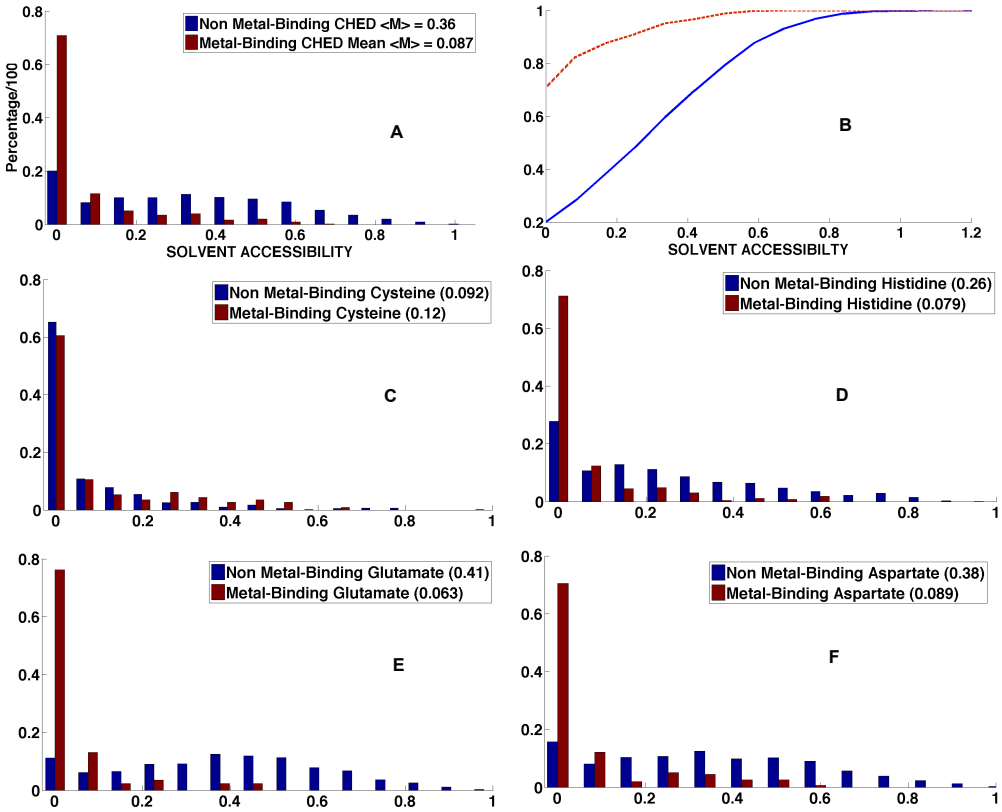


Figure 9. Comparison of the solvent accessibilities of metal binding and other residues.

A. Histograms of solvent accessibilities for CHED residues in the two groups**. B.** Cumulative distributions for two groups of CHED residues, indicating the strong propensity of metal-binding CHED residues to be buried. **C-F**. Same as panel **A**, for each of the CHED residues. Solvent accessibility was calculated using the constrained Voronoi procedure described in McConkey et al. (2002), and Eyal et al. (2004). The numbers in parenthesis on the legend show average solvent accessibility.

Thus we see a predisposition of metal-coordinating histidines, glutamates and aspartates to occupy regions with limited, if any, exposure to solvent. This property, along with the low

mobility noted above, can be used as 'features' in facilitating the identification of potentially metal-coordinating HED residues in putative metal-binding proteins.

2.4.6 Metal-binding sites have enhanced signal propagation properties

Chennubhotla and Bahar (2007) have introduced a Markov Model for describing the stochastics of signal transmission in proteins modeled as networks of nodes and springs. Two quantities are defined as metrics of communication propensities: (1) hitting time, *H(i,j),* which represents the number of steps (network edges) involved in transferring a signal from residue *j* to residue *i*, averaged over all possible paths connecting these two endpoints, and (2) commute time, *C(i,j) = H(i,j) + H(j,i).* The former depends on the direction of signal flow; the latter is independent of the direction. These two quantities are conveniently represented in terms of 2-dimensional maps, representative of the communication efficiency of all pairs of amino acids in a given protein. Notably, these two graph-theoretic concepts have been shown to correlate with physics-based properties obtainable by the GNM: commute time is directly proportional to MSFs in inter-residue distances (Chennubhotla & Bahar, 2007), residue pairs subject to small distance fluctuations being able to efficiently communicate. Hitting time, on the other hand, may be conveniently expressed in terms of the MSFs and cross-correlations in the positions of residues (see *Methods*).

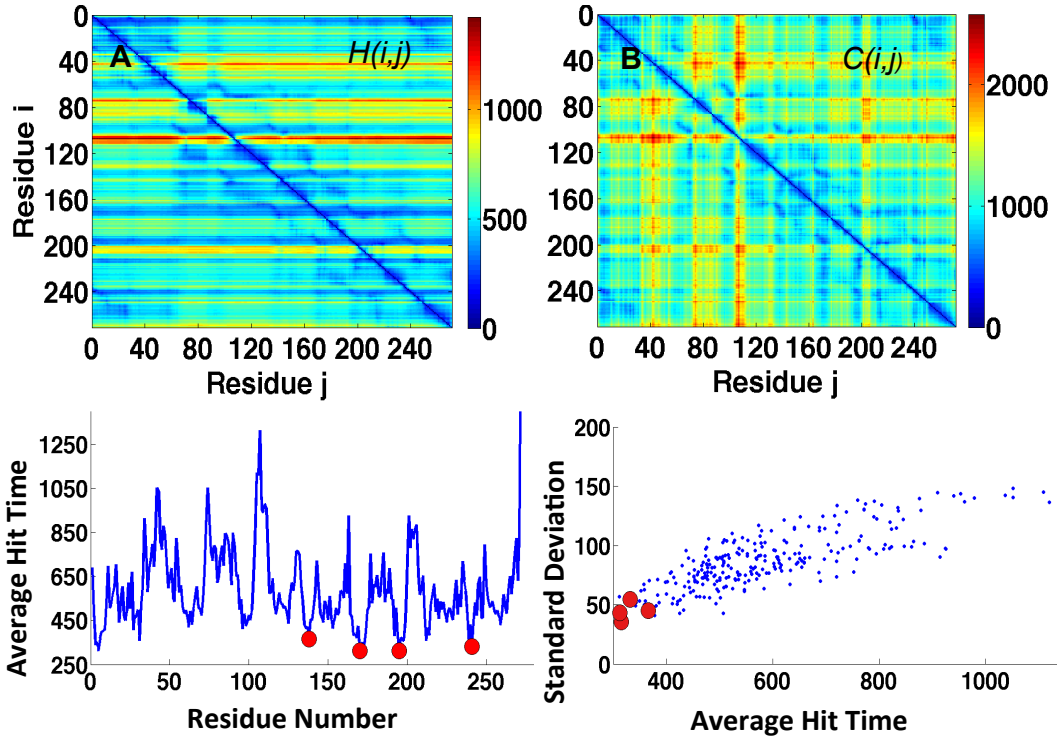


Figure 10. Signal propagation properties illustrated for a Zn^{2+} binding protein with endonuclease fold.

A. Hitting time *H(i,j)* map as a metric of the average duration of time, or number of steps, required to transmit signals from residue *i* to residue *j*, predicted by a Markovian model of communication (see Eq. 5). Blue and red regions correspond to shortest and longest hit times, respectively. **B.** Commute times *C(i,j) = H(i,j) + H(j,i)* (see Eq. 6). **C.** Average hit time *vs* residue number, evaluated from the average of *H(i,j)* over all starting points *j*. Red markers show the metal binding residues. Almost all of them occur at the minima of the curve, pointing to the efficient signal transmission properties of metal-binding sites. **D.** Average hit time *vs* its variance for each residue. The residues that participate in metal binding (shown in red markers) exhibit small average hit time along with small standard deviations.

Calculations performed for metal-binding proteins revealed that metal-binding sites possess uniquely efficient communication propensities. For illustrative purposes, we display in **Figure 10**, the types of the results for a zinc-binding protein (PDB ID: 1I6N). The top two maps describe the hitting times (**A**) and commute times (**B**) evaluated for all pairs. Notably, residues seem to have more or less uniform signal sending properties while they differ in their ability to

receive signals, as evidenced from the insensitivity of $H(i,j)$ to residue j. Panel **C** displays the average receiving properties of residues, found from $\langle H(i) \rangle = \mathbf{S}_j H(i,j)/N$. The red circles indicate metal-binding residues. Clearly, these residues have very low $\langle H(i) \rangle$ values, i.e. they are distinguished by their fast communication with all other residues. In addition to the mean hit times per residue, we also calculated the corresponding variance. The plot of the average hit time *vs* its standard deviation (or variance) for each residue in panel **D** clearly shows that metal-binding residues have minimal hit times and minimal variance, i.e. they are "fast and precise" in so far as their signal transmission properties are concerned. Note that similar features were observed for catalytic residues in our previous work (Chennubhotla and Bahar, 2007), suggesting that protein structures are designed to position their functional residues at key sites enabling cooperative response.

The results for all 145 metal-binding proteins are presented in **Figure 11**. As in **Figure 10D**, the points represent the mean (abscissa) and variance (ordinate) of the hit times $\langle H(i) \rangle$ evaluated for each residue in these proteins. In order to be able to display the results for all proteins on the same plot, hit times have been normalized with respect to the cumulative degree of the proteins (see *Methods*). Black dots correspond to CHED residues involved in metal binding (panel **A**) and other CHED residues (panel **B**). The comparison of the dispersion of black dots in the two panels demonstrates that metal-binding CHED (panel **A**) exhibit minimal hit time and variance compared to their non-metal-binding counterparts (panel **B**). We notice, however, two clusters of outliers in panel **A**. The former (enclosed by a red circle) refers to a Ni^{++}-binding transcription factor (PDB ID: 1B9N), also noted in Dataset I to be an outlier. The *apo* form of this protein has been resolved in the presence of molybdate ion, which elicited a significant change in its quaternary structure (Gourley et al., 2001), hence the difference in the global

dynamics of the *apo* and *holo* forms. The 2nd cluster (enclosed in the blue circle) in **Figure 11** panel **A** refers to a DNA-binding protein that might undergo atypical conformational changes upon DNA binding.

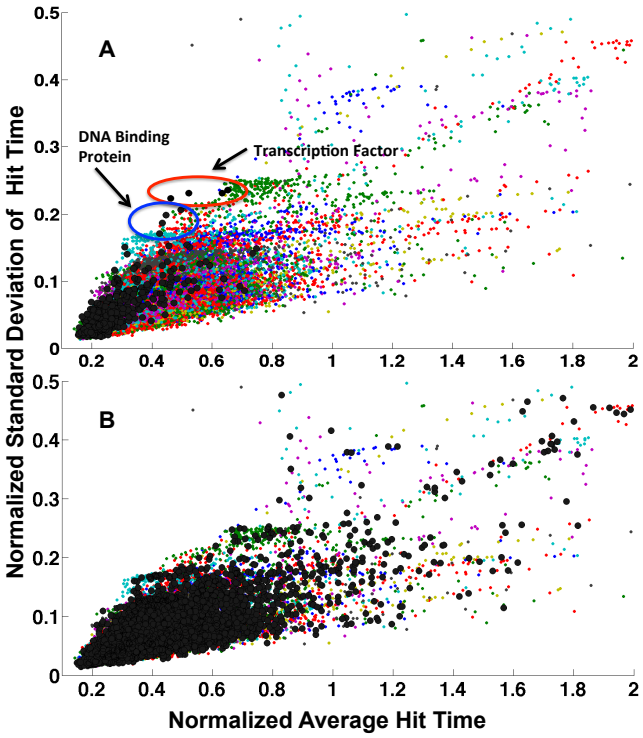


Figure 11. Comparison of the communication propensities of metal-binding and non metal-binding CHED residues.

Results are displayed for all residues in 145 *holo* metal-binding proteins. The abscissa shows the average hit-time, and the ordinate shows the variance in the hit-times, both quantities being normalized with respect to the cumulative degree of the proteins. **A.** The black markers show the position of metal-binding CHED residues; colored points refer to all other residues. **B** The black markers show the position of non metal-binding CHED. A few outliers are highlighted in panel **A**.

The above results point to the fast and effective communication property of metal binding residues and their surroundings. Thus it appears that based on the inherent network topology, proteins are intrinsically wired such that metal-binding residues are optimally positioned to enable efficient communication.

2.4.7 Insights into *de novo* design of metal-binding proteins

Assessment of collective motions intrinsically accessible to a given architecture can assist in designing proteins with suitable dynamics. Metal-binding residues emerge here as efficient signal propagators based on equilibrium state fluctuations available to the protein and show specific and fast communication patterns. These properties provide meaningful insights into the architectural design of metal-binding sites. These sites ought to be co-localized or near-neighboring with global hinges; and they should include charged and polar residues (e.g. CHED residues) that are buried, such that they will effectively ligate the metal and mediate the concerted motions of the surrounding structural elements, or the signal propagation between them.

To test the validity of this conjecture, we examined two *de novo* designed metal-binding proteins (**Figure 12**): a cobalt-bound metalloprotein in the form of a four-helical bundle (PDB ID: 1OVU) (Geremia et al., 2005) and a dimetal binding protein with Zn^{2+} (PDB ID: 2HZ8) (Calhoun et al., 2008). The global mode shape predicted by the GNM (panel A) as well as signal propagation properties reflected by the mean value and covariance of hitting times for each residue (panel B) clearly show that the metal-binding residues occupy key mechanical positions (minima in global mode) and have fast and precise communication capacities. Thus such inherent properties are also present, perhaps inadvertently, in the design of these proteins.

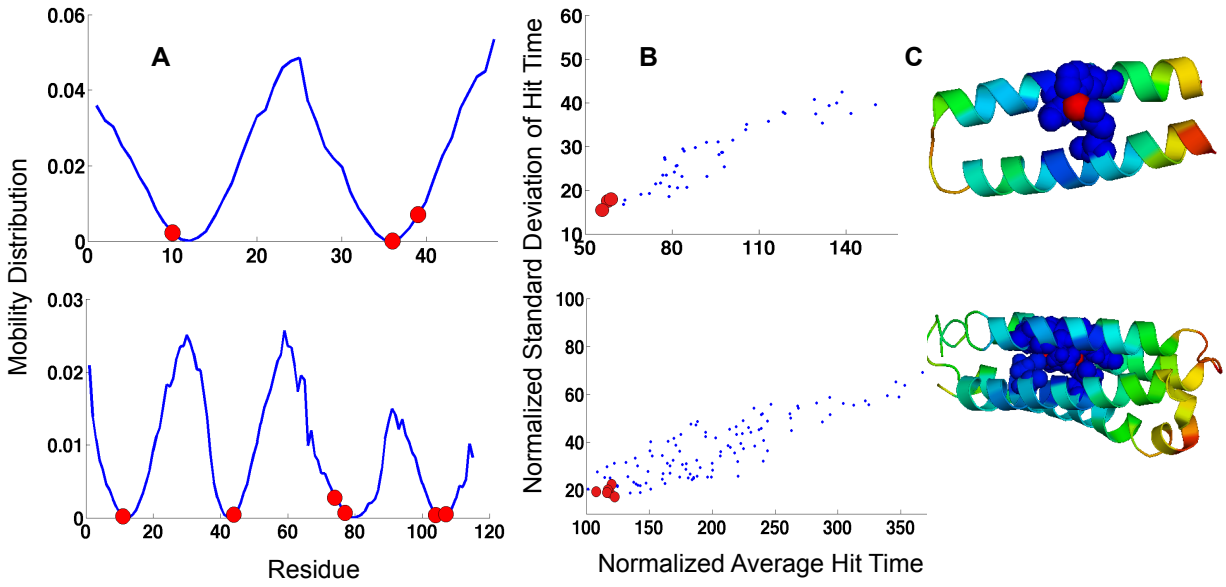


Figure 12. Global dynamics and signal transduction properties of metal-binding *de novo* designed proteins.

A. Global fluctuation profiles predicted by the GNM for 1OVU – a four-helix bundle binding Co^{2+} (*top*) and 1HZ8 – a dimetal Zn^{2+} binding protein (*bottom*). Metal binding sites are shown by the red markers. **B.** Average hitting times vs their covariance plot, for all residues in these two cases. Metal-binding residues (red circles) lie at the shortest hitting times and smallest variance region. C. Ribbon diagram color-coded by signal propagation properties with the blue regions indicating fast and efficient communication, and red regions the poorest communication properties. Metal-binding residues are shown by space-filling representation, and metals are shown by the red spheres. The metal binding sites were identified using Ligand Explorer in the PDB.

2.4.8 Conclusion

With increasing structural data on metal-binding proteins, we are now in a position to gain insights into the structural and dynamic features of metal-binding sites, and their significance with regard to the catalytic, transport, and signaling properties of the metal-binding proteins. Our results show that metal-binding sites might have an inherent preference to undergo minimal fluctuations in their positions, occupy central/buried positions despite being polar or charged, and possess unique signal transduction properties. These three properties are not necessarily

independent: more buried residues usually tend to have more restricted mobilities, and their tight packing confers efficient signal propagation properties. The fluctuations and signaling properties derived here are both based on network models: GNM for collective dynamics, and Markovian network model for allosteric communication. As described in the methodology, the residue fluctuations derived from the GNM relate to commute/hit times. Notably, the distinctive behavior of metal-binding His, Asp and Glu becomes more prominent when their signaling properties are examined, suggesting that these sites might be evolutionary selected to optimize the allosteric communication across the protein.

The study provides us with insights into simple design principles: the protein architecture uniquely encodes an ensemble of equilibrium motions, some being more probable than others. Functional residues/sites are usually implicated in some major way (e.g. hinge-bending, redistribution of salt bridges, conformational switches) in the softest motions, which are readily triggered by external perturbations (e.g. ligand binding). Metal-binding residues are indeed shown here to be located at/near such key mechanical positions (global hinge centers) to readily elicit cooperative responses.

The above arguments are exclusively based on topological properties of network models representative of protein structures. As such, they provide information on purely entropic driving forces. The results obtained here suggest that the entropic driving forces inherent to the geometry/architecture of the metal-binding proteins ascribe efficient mechanical and signal transduction properties to metal binding sites.

A review of ENM and Markovian signal propagation can be found in Eyal, Dutta and Bahar 2012 (Eyal et al., 2011) and the study on metal-binding proteins is published in Dutta and Bahar 2011 (Dutta and Bahar, 2010).

Next, we apply the methods described above in the following two chapters to analyze various aspects of the iGluR NTD dynamics (Chapter 3) and signaling propensities in the intact AMPAR (Chapter 4).

3.0 COMPARATIVE DYNAMICS OF GLUTAMATE RECEPTOR N-TERMINAL DOMAINS

The iGluR NTD comprises the most distal portion of the receptor and is believed to interact with presynaptic components and secreted factors including pentraxins in AMPARs (Hansen et al., 2010). The NTD, is structurally related to bacterial leucine-binding protein (Trakhanov et al., 2005), and also closely resembles the ligand binding cores (LBCs) of natriuretic peptide receptors (He et al., 2005) and type-C G-protein coupled receptors (GPCRs), including the type B g-aminobutyric receptor (GABABR) and the metabotropic glutamate receptors (mGluR1-8) (Pin et al., 2003). In mGluRs, glutamate binding within the NTD cleft triggers a ~ 30˚ interlobe closure motion and a rearrangement of dimer interface, which initiates G-protein signaling (Kunishima et al., 2000; Tsuchiya et al., 2002). The loose packing of the dimer interface has been thought to be crucial in facilitating glutamate binding and downstream signal propagation. Among its homologous iGluR counterparts, a well-established allosteric potential is only known for NMDARs (**Figure 2B**, showing ligand-binding sites). This domain has been suggested to merely act as a rigid subunit assembly module in nonNMDARs (based on GluA2 and GluK2 structures). However, the assembly characteristics of the AMPAR NTDs were seen to have unexpected diversity with GluA2 and GluA3 lying at functional extremes (Rossmann et al., 2010). The GluA3 NTD featured the weakest homodimeric affinity in solution and harbored conspicuous sequence variations in the lower lobe (LL) interface (more polar residues in the LL).

This difference in behavior led our collaborators to pursue X-ray crystallographic studies of the remaining GluA NTDs, to better understand the biology of this domain. Their efforts have resulted in the crystallization of previously uncharacterized GluA3 and GluA4 structure, and a higher resolution structure of the GluA2, enabling us to study the structure and dynamics of the NTDs of the AMPAR subfamily. To better understand this discrepancy between iGluR NTD activities we set out to compare NTD motions between the AMPARs, NMDARs and its homologue mGluRs (mGluR1) at multiple levels.

Utilizing a combination of X-ray structural approaches and NMA-ENM and all-atom based molecular dynamics simulations, we describe the structural variations seen across all AMPARs and their repercussions on dynamics. We also characterized the currently elusive dynamics of the NMDAR NTD, a powerful allosteric modulator. Using ENMs we establish that AMPAR NTDs do have access to well-defined collective motions. Specifically, we detected counter-rotations and clamshell-like lobe-motions analogous to those seen in iGluR LBDs (Armstrong and Gouaux, 2000) and in mGluRs (Kunishima *et al.*, 2000). When comparing the intrinsic dynamics of AMPAR- and NMDAR NTDs, we see that the global modes of motion accessible to AMPAR NTD monomers are almost identical to those of NMDAR NTDs, albeit more restricted upon dimerization. Moreover, AMPAR NTDs possess the ability to readily reconfigure into conformers comparable to those of NMDAR NTDs.

All-atom molecular dynamics (MD) simulations reveal clear differences in dimeric interactions and their time evolution between the AMPAR paralogs, particularly at the LL interface, which unzipps and ruptures in GluA3. Further MD simulations of mutants designed to either destabilize or strengthen the LL intersubunit interactions further imply a key role played by LL interface contacts in mediating global motions, which might be transmitted to the LBD.

Combined with MD and ANM results, we determine the mechanisms of global and local motions favoured by the iGluR NTD architecture, identify critical residues facilitating these motions, and reveal a mechanistic link between LL interface stability and cleft movements, which vary among subfamily members. Together, these data reveal an unexpected parallel between AMPAR and NMDAR NTDs, and provide a first glimpse into the dynamic landscape of iGluR NTDs.

3.1 STRUCTURAL DETAILS OF NEWLY CRYSTALLIZED AMPAR NTDS

3.1.1 GluA3 structural features

A comparison between the structures of GluA2 and GluA3 shows that the most striking feature observed in GluA3 structure is a repositioning of the LLs, which in GluA3 dimer I (3O21-*CD*) are widely separated, ~ 7 Å away from the axis of 2-fold symmetry relative to GluA2 (**Figure 13A-B**). The LL arrangement observed in the GluA3 NTD bears a striking resemblance to mGluR1- and the natriuretic peptide receptor LBCs, where signaling via flexible LLs is well established (Kunishima et al., 2000; Tsuchiya et al., 2002). In fact GluA3 and mGluR1 show a very similar degree of LL separation. Thus, unlike GluA2, in GluA3 the LLs are not constrained by dimeric packing but are free to move and may propagate signal. It is worth pointing out that ligand-independent clamshell motions have been deduced from experimental data in NMDARs (Gielen et al., 2009); a related scenario can be expected in GluA3 (see below). The structure also provides an immediate explanation for the relatively low GluA3 NTD dimer affinity measured in solution (Rossmann et al., 2010).

A closer examination of the LL interface reveals that, contrary to GluA2, the GluA3 LL interface is largely polar, which was not anticipated previously from sequence alignments (Jin et al., 2009). In particular, Arg163 and Arg184 project towards the interface (**Figure 13C**) generating positive electrostatic potential; charge repulsion presumably contributes to the increased lobe separation seen in dimer I. Arg163 is replaced by hydrophobic residues in the other AMPAR subunits - in GluA2 Ile157 takes its place and engages Ala148 of the opposite protomer in hydrophobic contacts (**Figure 13D**). GluK1-3 kainate receptors also harbor Arg at this position; however, the positive charge is shielded effectively by Glu186 and Glu192 (**Figure 1D**) (Kumar et al., 2009); interestingly, LLs in kainate receptors are well conserved which is in apparent contrast to AMPARs (**Figure 1D**). We conclude that the previously described 'locked' GluA2 dimer, which is also seen in the GluK2 kainate receptor (Jin et al., 2009; Kumar et al., 2009), is not universally found in all nonNMDARs.

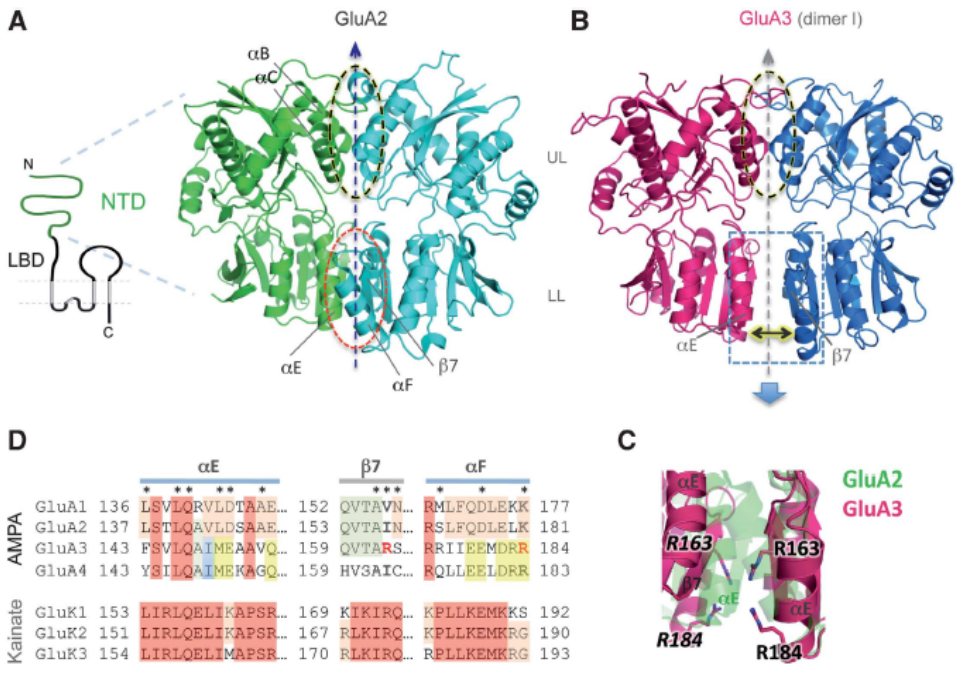


Figure 13. GluA2 and GluA3 NTDs differ structurally.

A. *Left:* Topology of an iGluR subunit. The NTD segment is denoted as a green curve and the transmembrane segments as grey columns. *Right:* Structure of the bipartite GluA2 NTD dimer (PDB 3HSY). The two chains/protomers are coloured green and cyan. Upper and lower lobes (UL, LL) are denoted and their respective interprotomer interfaces are circled. Secondary structural elements contributing to the LL interface are labeled. **B** Structure of the GluA3 NTD (3O21 - *CD*), with the two protomers coloured red and blue. The UL dimer interface analogous to GluA2 is circled, and the LL interface is shown by a box and an arrow indicating the increased space between the LLs, compared to GluA2. Segments homologous to the GluA2 LL interface segments (from A) are labeled. (C) Lower lobe packing markedly differs between GluA2 and GluA3 NTDs. The lower lobe interface of GluA2 (green) and GluA3 (red) are shown after aligning common secondary structure segments. Note the significantly closer packing of the GluA2 LL interface. Also shown are arginines from GluA3 that project into the interface; this unfavorable electrostatic interaction may contribute to the increased inter-lobe distance. (D) Sequence conservation in the NTD LL of the AMPA and kainate subfamilies. Different background colours indicate different conservation patterns; e.g. conserved sites (columns) within a subfamily are coloured red. Residues that project across the interface are denoted with asterisks (*). Note the markedly different conservation pattern within the kainate subfamily, whose LL interface is overall more conserved.

3.1.2 GluA4 structural features

GluA4 resembles other AMPAR (GluA1-A3) NTDs with root-mean-square deviation (RMSD) values of ~1.2Å when superimposing individual NTD protomers, and up to 4.1Å when superimposing the NTD dimers (**Figure 14A** for monomers and **Table 3** for dimers). As in GluA1 and GluA3 (Sukumaran *et al.*, 2011)(Yao *et al.*, 2011), no ligand density was evident in the inter-lobe binding cleft, and lobe opening angles were similar between GluA1-GluA4.

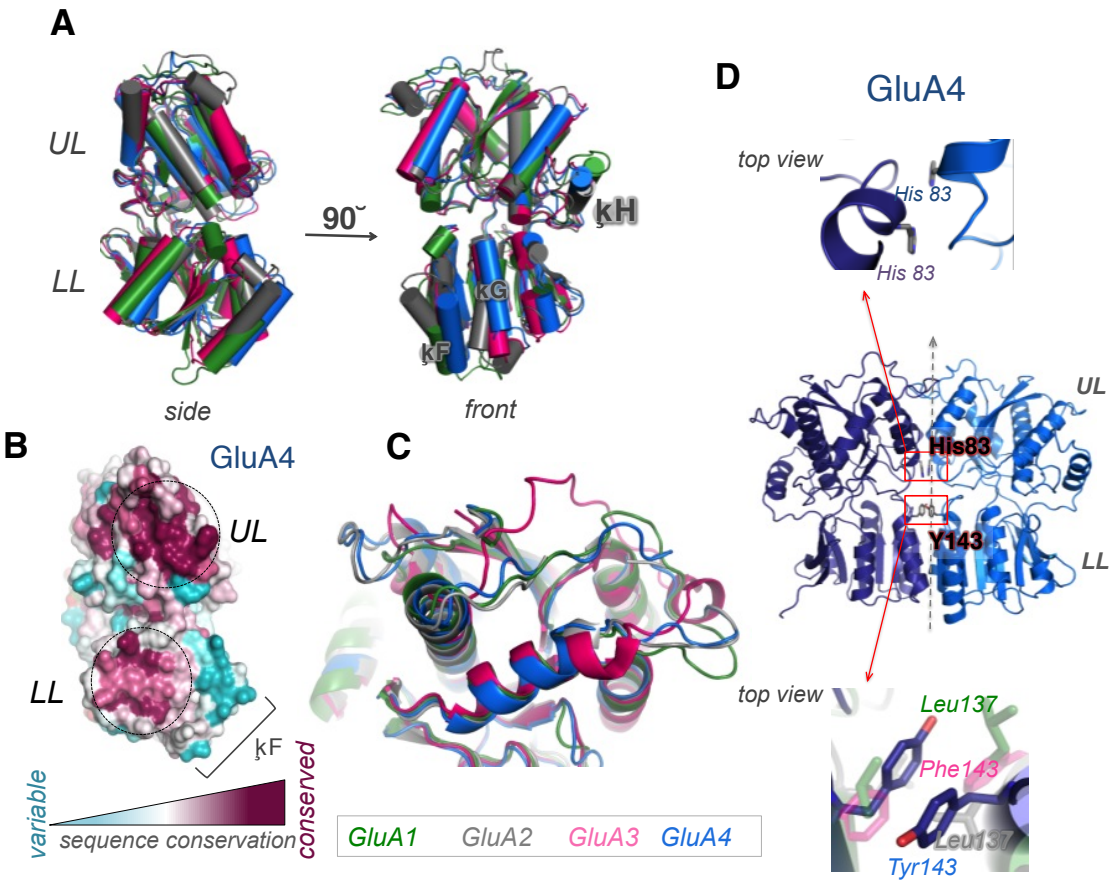


Figure 14. Comparative structural analysis of AMPAR NTDs.

(A) Structural superposition of all four AMPAR NTD monomers shown in side and front views (GluA1: PDB 3SAJ, *green*; GluA2: PDB 3HSY, *grey*; GluA3: PDB 3O21, *red*; GluA4 (new structure), *blue*). (B) Dimeric interface of GluA4, colored by position-specific patterns of conservation (generated by the ConSurf package (Ashkenazy H et al, 2010) and a manually curated alignment of 142 AMPAR paralogs. The UL interface is more conserved than the LL interface (C) Focuses on the superposition of the top loops that show some degree of variability in all AMPAR NTD paralogs. (D) Shows the difference in interfacial packing of two critical residues in GluA4 compared to other AMPAR counterparts.

The UL interface is the most highly conserved region between AMPAR NTD paralogs (**Figure 14B**). Packing along the GluA4 upper lobe (UL) dimer interface is indeed comparable to that of other AMPAR NTD subtypes, with variations mostly in positioning of the top/selectivity loops (**Figures 15B** and **Figure14C**). An interesting difference is His83 projecting from the base of α-helix C (αC) across the GluA4 dimer interface; this residue is replaced by Asn in GluA1-3

(**Figures 15B** and **Figure14D**). The LL interface is more variable in primary sequence and in packing geometry between GluA1-4 (**Figure 15B**). Contacts across the LL interface of GluA4 are comparable to GluA2 (Protein Databank (PDB) structure 3HSY), to GluA3 dimer *BD* (PDB 3P3W; italicized letters indicate chain identifiers from PDB structures, i.e. protomers *B* and *D* from 3P3W) and to GluA1 dimer *AC* (PDB 3SAJ) (Yao *et al.*, 2011). However, in GluA3 dimers *CD* (PDB 3O21) and *AC* (PDB 3P3W), the inter-LL distance is wider (Sukumaran *et al.*, 2011) (**Figure 15B**). Also in GluA1 (dimer *BD*; PDB 3SAJ), the LLs are packed more loosely partly due to rotations of Leu137 away from the interface. At the equivalent position, GluA4-selective Tyr143 side chains stack across the interface, separated by ~3Å (**Figure 14D**). Thus, UL interface contacts are mostly conserved in AMPAR NTDs whereas LL packing is diverse and subunit selective. The looser LL contacts in GluA1 and GluA3 correlate with reduced NTD dimer stabilities (Rossmann *et al.*, 2011).

Table 3. RMSD values (Å) between different AMPAR NTD dimeric structures

RMSD	GluA1-*AC*	GluA1-*BD*	GluA2-*AB*	GluA3-*AB*	GluA3-*CD*	GluA4-*BA*
GluA1-*AC*	-	0.815	1.019	3.382	2.901	1.264
GluA1-*BD*		-	1.067	3.323	2.864	1.282
GluA2-*AB*			-	3.418	3.333	0.991
GluA3-*AB*				-	4.094	3.163
GluA3-*CD*					-	2.430
GluA4-*BA*						-

To quantify contacts across the structurally variable LL interfaces, we calculated local atomic contact density (LD) indices, a measure for packing density across interfaces (Bahadur *et al.*,

2004; Sukumaran *et al.*, 2011) (**Figure 15C**). GluA4 exhibits interface characteristics similar to

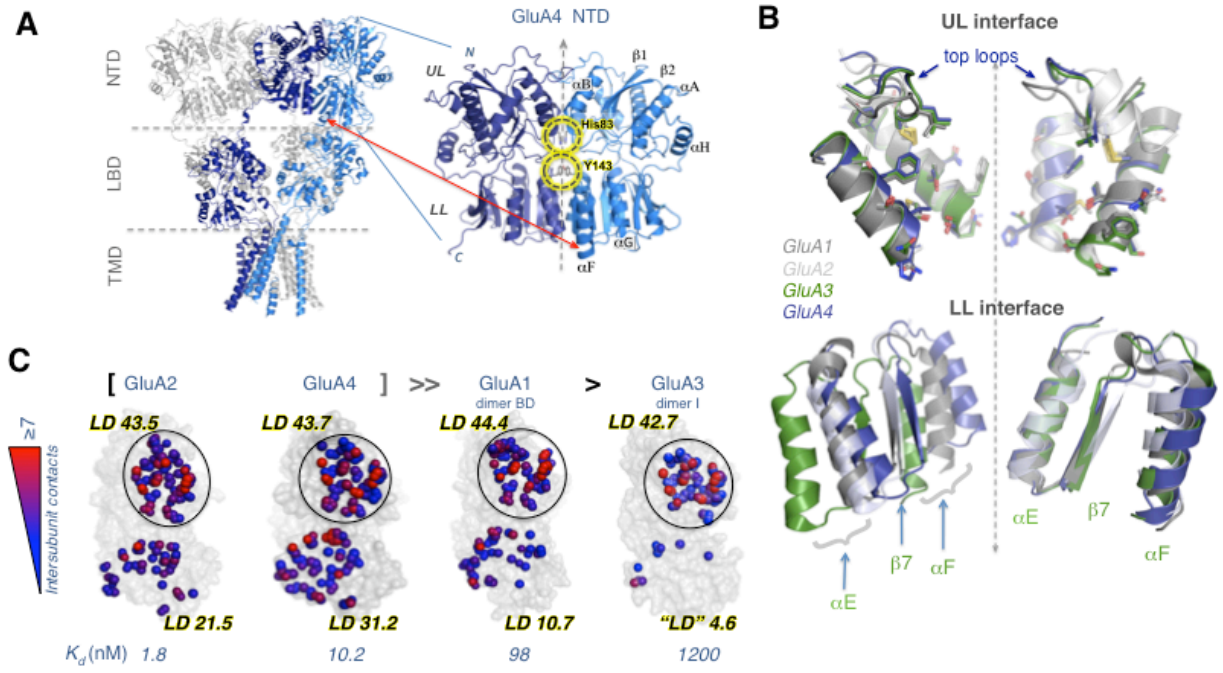


Figure 15. Structure of the GluA4 NTD facilitates a comparative structural analysis.

(A) Intact structure of GluA2 AMPAR (*left*) displaying the spatial arrangements of four subunits (two shown in *gray*, and the others in *blue* and *dark blue*) that span the three domains (NTD, LBD and TMD). The location of the NTD dimer resolved for GluA4 is enlarged. Subunits are symmetrically positioned, each consisting of an upper lobe (UL) and a lower lobe (LL); secondary structural features (helices αA, αB, αE, αF, αG, and αH and strands β1 and β2) are labeled. Interfacial interactions are highlighted. **(B)** UL dimerization interfaces of GluA1-4 are largely conserved but LL packing shows heterogeneity. UL interfaces of GluA1-2 (*greys*), GluA3 (*green*) and GluA4 (*blue*) have been artificially separated to show the structural conservation and orientations of key residues (shown in stick) making contacts across the interface. Two-fold axis of symmetry is shown as a dashed line. Superposition of LL shows distinct differences in interface packing that is most prominent in GluA3. **(C)** Intersubunit contacts at the UL and LL interfaces of GluA1-4 NTDs. Atoms making interfacial contacts within 4.5Å are shown as spheres and colored from *blue* (1 contact) to *red* (≥ 7 contacts). Calculated local contact density (LD) indices and empirically measured dimer dissociation constants (K_d) are also shown. The four NTDs are ranked by their homodimerization affinity. Note the LL interface is highly variable between AMPAR paralogs, whereas the UL interface is largely invariant (**Figure 14**).

those of GluA2 (UL interface LD: 43.7, LL interface LD: 37.2). GluA1 shows high contact density in the UL (LDs of 40.8 and 44.4 in both dimers *AC* and *BD*), characteristic of tight,

biologically relevant interfaces (Bahadur *et al.*, 2004). However, the LL interface shows variability: GluA1 dimer *AC* is similar to those of GluA2 and GluA4, whereas *BD* is similar to GluA3 (*CD*; PDB 3O21). Again, these structural features agree with measurements of AMPAR NTD homodimer stabilities by analytical ultracentrifugation (Rossmann *et al.*, 2011), where GluA3 exhibited the weakest affinity (K_d ~1μM), followed by the intermediate GluA1 (K_d ~100nM), likely reflecting their LL separations and multiplicity of dimeric forms in crystal structures. GluA2 and GluA4 featured K_d values between 2 and 10 nM, consistent with tighter LL packing (Zhao *et al.,* 2012). Thus, the greatest structural variability between AMPAR NTDs is at the LL dimer interface; looser LL packing is expected to increase NTD inter-protomer mobility.

3.2 COMPARATIVE ANM ANALYSIS REVEALS GLOBAL MOTIONS SHARED BY AMPAR NTDS

To compare the structural dynamics between AMPAR NTDs and to contrast those to the allosterically active NMDAR NTDs, first we analyzed the collective motions of GluA1-4 using the ANM (Atilgan *et al.*, 2001; Eyal *et al.*, 2006). ANM is particularly suitable for a comparative assessment of global motions intrinsic to whole protein families (Bahar *et al.*, 2011). Global motions are those at the lowest frequency end of the mode spectrum, predicted by the ANM to be uniquely defined by the native fold. The lowest frequency mode, *mode 1*, represents a structural change (usually subunit/domain movements) along the softest/smallest ascent direction away from the original energy minimum (Bahar *et al.*, 2010).

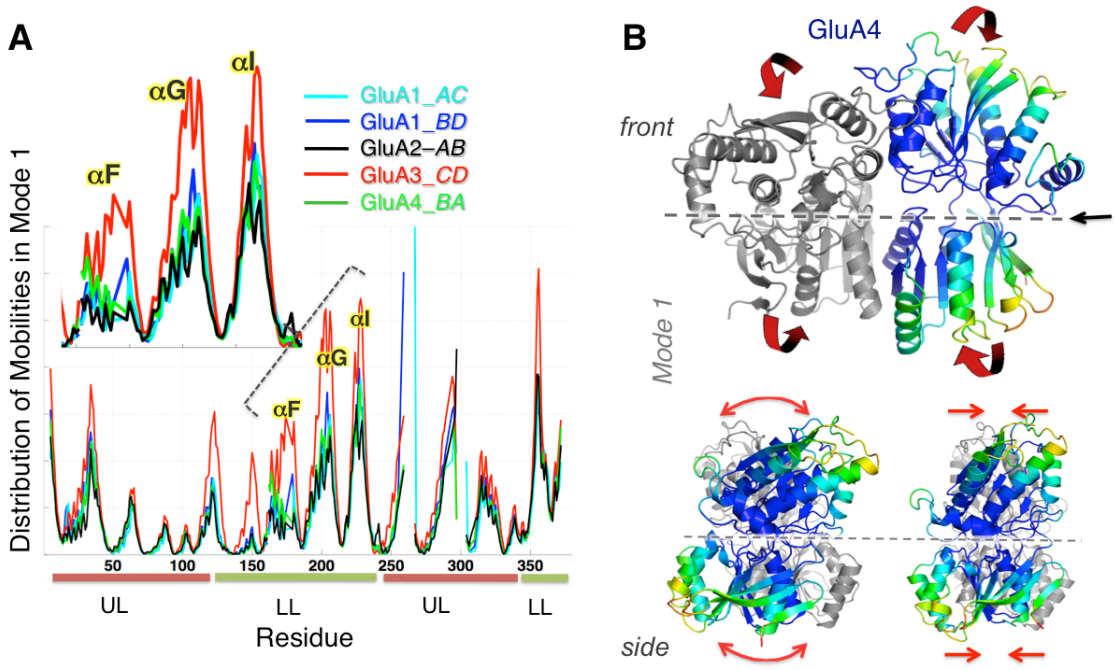


Figure 16. Global dynamics of GluA4 dimer in comparison to other AMPAR NTDs probed by ANM.

(A) Distribution of square displacements of residues in the most global (lowest frequency) mode intrinsically accessible to AMPAR NTD dimers (GluA1-*AC*, GluA1-*BD* (3SAJ), GluA2-*AB* (3HSY), GluA3-*CD* (3O21) and GluA4 (new dimeric structure). The four subtypes show similar profile (see the high correlations listed in **Table 4**), but different size motions (see **Table 5**). (**B**) Shared mechanism of global motion: counter-rotation of the two protomers (indicated by *red* arrows), depicted for GluA4 as a representative structure, from the front and side view. The diagram is color-coded from *red* (most mobile in *mode 1*) to *blue* (least mobile). The global mobility rank of the four AMPAR NTD dimers is GluA3-*CD* (0.110) > GluA1-*BD* (0.169) > GluA1-*AC* (0.184) ≈ GluA4-*BA* (0.187) > GluA2-*AB* (0.187). The numbers in parentheses indicate the global mode eigenvalues (see *Methods*).

ANM calculations performed for GluA1-4 NTD dimers showed that the AMPAR subfamily members share a common mechanism of global reconfiguration with a correlation of 0.90 ± 0.04 (**Figure 16** and **Table 4A**): torsional counter-rotation of the two protomers, as previously noted for GluA2 and GluA3 (Sukumaran *et al.*, 2011) and extended here to the entire subfamily (**Figure 16A-B**). The four subunits exhibit similar global mode shapes (distribution of mobilities, **Figure 16A**); their motion amplitudes (peak heights), however, show variations, with

GluA3 exhibiting higher mobilities (**Table 5**), particularly in the LL (residues 120-225; inset of **Figure 16A**). These data, i.e., the flexibility between paralogs (**Figure 16A** inset), directly correlate with experimental measurements of AMPAR NTD dimer K_d's (**Figure 15C**) (Rossmann *et al.*, 2011).

Table 4. Correlations* between the global motions favored by iGluR subfamily NTDs

A. Correlations between the first global mode of AMPAR NTD dimer subtypes						
		AMPAR NTD Dimers				
		GluA1-*AC*	GluA1-*BD*	GluA2-*AB*	GluA3-*CD*	GluA4-*BA*
AMPAR NTD Dimers	GluA1-*AC*	-	0.983	0.897	0.876	0.922
	GluA1-*BD*		-	0.916	0.912	0.939
	GluA2-*AB*			-	0.874	0.926
	GluA3-*CD*				-	0.926
	GluA4-*BA*					-

B. Correlations between global modes of AMPAR and NMDAR NTD protomers					
		NMDAR NTD Monomers			
		GluN1-*A* (Mode 1)	GluN1-*A* (Mode 2)	GluN2B-*A* (Mode 1)	GluN2B-*A* (Mode 2)
AMPAR NTD Monomers	GluA1-*A*	0.757	0.756	0.731	0.757
	GluA2-*B*	0.765	0.672	0.816	0.628
	GluA3-*C*	0.734	0.713	0.723	0.675
	GluA4-*A*	0.841	0.823	0.796	0.824

(*) Correlations are evaluated as the cosines between the 3N-dimensional eigenvectors corresponding to the softest modes of collective motions predicted by the ANM

Table 5. Relative sizes* of NMDAR and AMPAR NTD global motions

AMPAR NTD Dimers						
GluN1 *AB*	GluN2B-N1 *CD*	GluA1 *AC*	GluA1 *BD*	GluA2 *AB*	GluA3 *CD*	GluA4 *BA*
Mode 1						
1	0.87	0.25	0.27	0.24	0.42	0.25
AMPAR -NMDAR NTD Monomers						
GluN1-*A*	GluN2B-*A*	GluA1-*A*	GluA2-*B*	GluA3-*C*	GluA4-*B*	
Mode 1						
1	0.58	0.54	0.52	0.59	0.52	
Mode 2						
1	0.48	0.53	0.45	0.48	0.43	

(*) based on the eigenvalues of the softest modes predicted by the ANM, normalized with respect to GluN1-A in case of monomers and GluN1-AB in case of dimers

3.3 NMDAR INTRINSIC DYNAMICS SUPPORTED BY ANM AND MD

NMDAR NTDs allosterically modulate NMDAR ion channel function, triggered by small molecule ligands and Zn^{2+} ions that bind the dimer interface and cleft region between lobes, respectively. However, current NMDAR NTD structures are similar with regard to cleft-opening angle (Karakas *et al.*, 2009) and protomer conformation upon ligand binding (Karakas *et al.*, 2011). Thus, the motions underlying NMDAR NTD allostery are unknown.

Our ANM analysis of the global dynamics of the NMDAR NTD, performed for the NR2B subunit, revealed a global twist of the LLs towards the dimer interface accompanied by cleft opening, whereas LL twist motion in the opposite direction induced cleft closure, highlighting the classic clamshell-like motion (**Figure 17A**). Full-atom molecular dynamic (MD) simulations performed for the same subunit, in the *apo* (NMDA1) and Zn^{2+}-bound (NMDA2)

states also revealed an overall rigidification accompanied by cleft-angle closure in the presence of Zn^{2+}; (**Figure 17B**, *pink* curve), whereas in the absence of Zn^{2+}, cleft opening was observed (*blue* curve). The cleft angle was monitored based on the relative positions of L124 (UL), S149 (cleft) and I257 (LL) α-carbons. The apo form thus stabilizes a more open conformation by at least 17˚ compared to the Zn^{2+}-bound form (**Figure 17B**, inset). Strikingly, the same type of structural change is predicted by the ANM mode 2 for NR2B (**Figure 17A**). Thus both ANM and MD support a classic periplasmic-binding protein mode of ligand recognition upon Zn^{2+} binding.

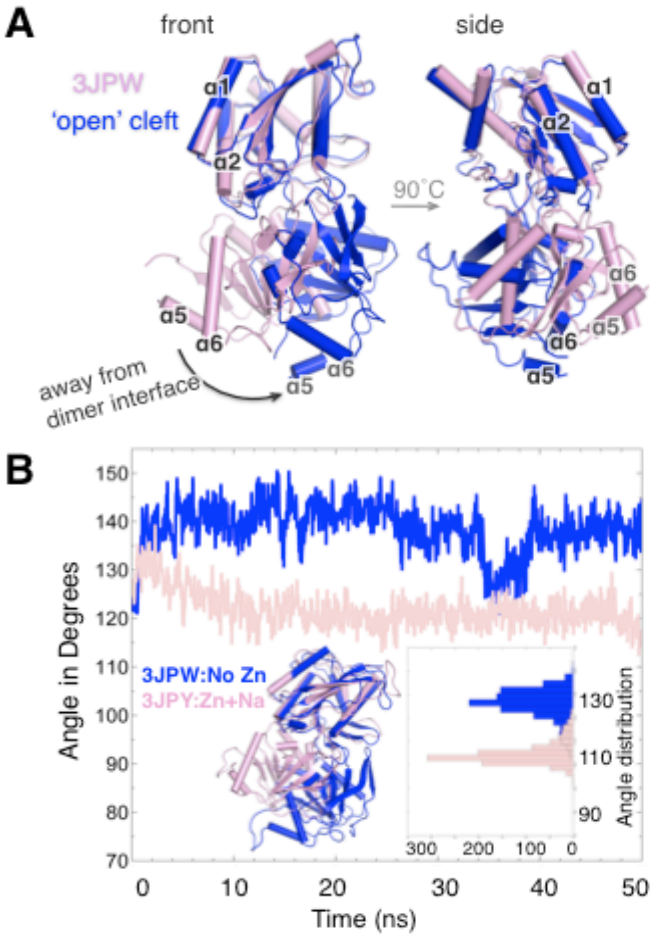


Figure 17. Intrinsic ability of NMDAR NTD to undergo cleft motions.

(A) Deformation of NR2B subunit (PDB 3JPW, *pink*) along ANM *mode 2*, leads to opening up of the cleft (*blue*). (B) The time evolution of the cleft angle observed in the MD runs NMDA1 (*pink*, in presence of Zn^{2+}) and NMDA2

(*blue*, in absence of Zn^{2+}). The cartoon in the inset is the superposition of 50ns snapshots from NMDA1 and NMDA2. It illustrates the opening of the cleft in the simulation performed without Zn^{2+} similar to the global reconfiguration predicted by the ANM for 3JPW in A. The histograms in the inset are of the distribution of the angles sampled by NR2B in the two simulations: the average angle is 138˚ in NMDA1, and 121˚ in NMDA2.

3.4 COMPARISON OF NMDAR AND AMPAR DYNAMICS

3.4.1 Similarity between the intrinsic dynamics of NMDAR and AMPAR NTD protomers

We next compared global dynamics between AMPAR and NMDAR NTDs. Despite their distinctive structural features (Furukawa, 2012), the global modes between the two subfamilies are surprisingly preserved. **Figure 18** illustrates the results for AMPAR GluA2 and GluA3, and NMDAR GluN1 and GluN2B subunits. Two dominant modes of motion are observed: counter-rotation between the two lobes (*mode 1*; panels **A-B;**) and intra-lobe clamshell opening/closing (*mode 2*; panels **C-D**). The global modes of all AMPAR and NMDAR NTD monomers exhibited a high level of similarity, with correlation coefficients varying in the range 0.73 ± 0.11, highlighting the *universality* of the observed motions despite stark differences in tertiary and quaternary packing (**Table 4**). Especially, the clamshell-like motion seen in NR2B (**Figure 17**), which enables sampling of ligand unbound/bound conformations, is also preserved in AMPAR monomers, hinting at their ligand-binding potential.

While global mode shapes are similar between GluN1, GluN2B, GluA2 and GluA3 monomers (**Figure 18A-C**), the relative amplitudes are largest in GluN1, and smallest in GluA2. **Table 5** shows an overall ranking of GluN1 > GluA3 ~ GluN2B > GluA1 > GluA4 ~ GluA2 (stiffest) based on *mode 1* (see Methods), and a similar trend is observed in *mode 2*. This analysis reveals an unexpected difference between GluN1 and GluN2B. Importantly, these modes of

motions and intrinsic flexibility are largely conserved between NMDA- and AMPAR NTDs, as discussed further below.

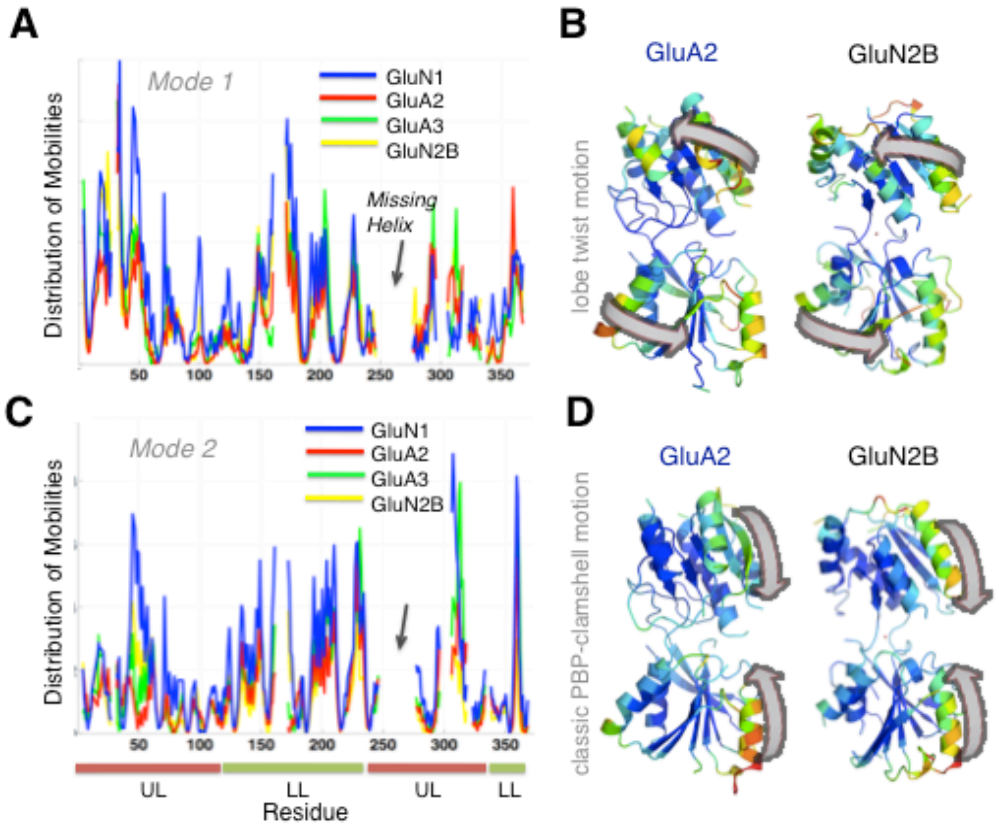


Figure 18. Comparing the global dynamics of NTD protomers resolved for AMPA and NMDA receptors.

(A) Comparison of the mobility profiles as driven by the lowest frequency (most cooperative) mode of motion accessible to GluN2B (3JPW), GluN1 (3Q41- *A*), GluA2 (3HSY-*B*) and GluA3 (3O21-*C*) NTD monomers. The abscissa in A is labelled according to residues in GluA2. (B) Ribbon-diagram of a representative AMPAR (GluA2) and an NMDAR (GluN2B) NTD monomer, colored by the mobility profile in *mode 1*. The arrows indicate the mechanism of motion (counter-rotation of the two lobes). (C) and (D) same as panels A and B, for ANM *mode 2*, a clamshell-like opening/closing of the two lobes.

3.4.2 Effect of dimeric packing on dynamics of AMPAR and NMDAR NTD monomers

Since AMPAR- but not NMDAR NTDs assemble into stable homodimers (Clayton *et al.*, 2009; Rossmann *et al.*, 2011), we next evaluated the changes in dynamics upon NTD dimerization, using a perturbation method (Zheng and Brooks, 2005; Ming and Wall, 2005b), which facilitates assessing the effect of environment on the dynamics of examined systems. Here, each monomer in the dimeric NTD of AMPAR (GluA2 and GluA3 homodimers) and NMDAR (N1-N2B heterodimer) is taken as the *system* and is analyzed in the context of the other monomer, which represents its *environment*. The dynamics in the presence of the environment is then compared to that of the system in isolation (i.e. the intrinsic dynamics of the monomers, presented above).

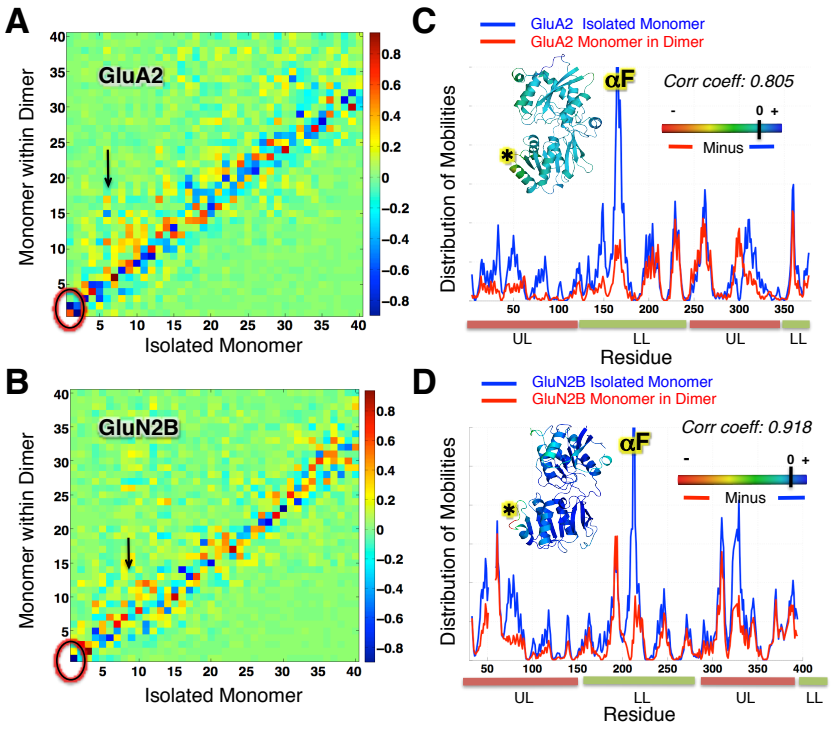


Figure 19. Effect of dimerization on the intrinsic dynamics of AMPAR and NMDAR NTD monomers.

(A) Correlation between top 40 modes accessible to GluA2 protomer in isolation (3HSY-*B*; *abscissa*) and in the dimer (3HSY; *ordinate*). Darkest *red* and *blue* regions refer to strongest correlations (see the scale on the right). Clamshell motions (monomer *mode 1*) are maintained in the dimer, but manifested by *mode 2* (circled region). (B) Same as A, for GluN2B (3QEL-*D*) monomer compared to GluN1/GluN2B heterodimer (3QEL); (C, D) Mobility profiles for GluA2 and GluN2B monomers in isolation and in the dimer, showing the suppression of mobilities (at the UL in particular) upon dimerization (see Figure S2 for GluA3 and GluN1). Insets show GluA2 and GluN2B monomers colored by their change in mobility upon dimerization, from most suppressed (*red*) to unaffected (*blue*).

The maps in **Figure 19A-B** display the correlations between the top-ranking 40 ANM modes predicted for the isolated and dimeric forms of the monomers of GluA2 and GluN2B. Highest correlations are indicated by correlation cosines (see *Methods*) close to ± 1 (colored *red/blue*), and lack of correlation by values approaching zero (*green*). The observed high correlations along the diagonal indicate that the dynamic character of the monomer is maintained in the dimer, with minor alterations (and occasional swaps in the order of mode).

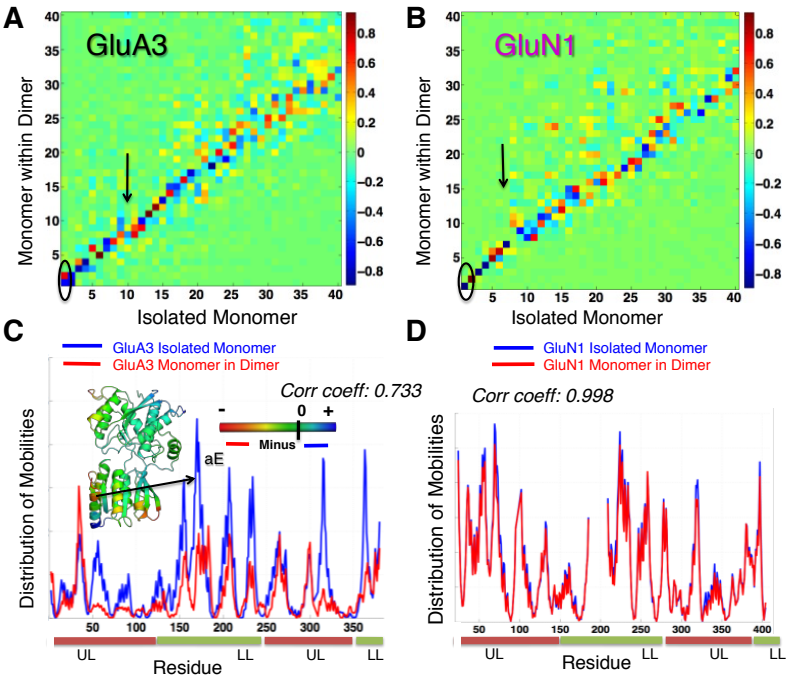


Figure 20. Effect of heterodimerization on NTD monomer dynamics of AMPA and NMDA receptors

(A) Correlation cosine between the top 40 eigenmodes accessible to the isolated GluA3 protomer (3O21-*C*; *abscissa*) and the same protomer in the dimer (3O21-*CD*; *ordinate*). The encircled region shows that isolated

monomer *mode 1* is accounted by the *2nd mode* in the dimer. **(B)** Same as **(A)**, for NMDA GluN1 (3QEL-*C*) isolated monomer compared to its behavior in the heterodimer (3QEL) **(C, D)** Mobility profiles for GluA3 and GluN1 monomers in isolation and in their respective dimers, showing the suppression of mobilities upon dimerization in both AMPAR GluA3, but not in GluN1 that retains its flexibility on dimerization. The inset in **(C)** shows GluA3 monomer colored by difference in mobility i.e., red curve – blue curve. The color bar indicates that red implies + values or regions where monomer motion is most suppressed in the monomer on dimerization and blue showing regions where motion is suppressed in the isolated form.

While the shapes of the global modes are closely maintained, the amplitudes of the motions exhibit a dependence on dimerization. One would expect the amplitudes of fluctuations to be depressed by inter-protomer interactions, especially at interface regions. This is the case for GluA2, GluA3 and GluN2B (**Figure 19C-D** and **Figure 20C**) where the protomer in the dimer exhibits lower mobility compared to the isolated monomer. The insets in **Figure 19C-D** show the ribbon diagrams of the GluA2 and GluN2B monomers respectively, colored by the difference in mobility between the monomer in the dimeric system and the isolated monomer. The region that shows the largest suppression is αF in the LL followed by the UL interface, while UL and LL cores remain unchanged. Dimerization has almost no effect on the mobility of GluN1, i.e., inter-protomer interactions do not obstruct the conformational flexibility of this NTD (**Figure 20D** and **Table 4**). Notably, the suppression of mobility in the αF helix region may have implications on the allosteric capacity of AMPAR NTDs.

3.4.3 NMDAR and AMPAR NTDs readily reconfigure along a single, global mode of motion

The observed difference in the size of global motions between GluN1 and AMPAR NTD protomers are likely due to their differences in dimeric packing. We next determined whether

dimer conformations are interconvertible between iGluR subfamilies. If high-energy barriers separate different dimeric forms and preclude structural rearrangement, their distinctive (non-convertible) inter-protomer packing would impact their dynamics, and the known allosteric capacity of the NMDAR NTDs could be attributed to higher conformational freedom. If however, the different structures are alternative forms readily accessible via soft modes of motions, this would imply that the seemingly less mobile AMPAR NTDs (such as GluA2 and GluA4) can access conformers with known allosteric potential (i.e. NR2B).

To make a quantitative assessment of the ease of transition between different NTD dimers, we examined the *overlap* (see *Methods*) between (i) structural difference, $\Delta\{\mathbf{R}\}_{S1\rightarrow S2} = \{\mathbf{R}_0\}_{S1} - \{\mathbf{R}_0\}_{S2}$, that is required for the transition from dimeric conformer 'S1' to conformer 'S2' (based on PDB coordinates, after optimal superimposition of the endpoints), and (ii) the soft modes of structural change favored by S1, as predicted by the ANM. A strikingly easy 'conversion' between AMPAR and NMDAR NTD conformers is evidenced by the high overlap between $\Delta\{\mathbf{R}\}_{S1\rightarrow S2}$ and *mode 1* predicted for S1. **Figure 21A** and **B** illustrate the results for GluN1→GluA2) and GluA3→GluN1-N2B, respectively. The former provides evidence for the ease of transition from NMDAR (GluN1) homodimer to the GluA2 homodimer; and the latter, from the GluA3 homodimer to the GluN1/GluN2B heterodimer (**Figure 21C-D**). This analysis underscores the significance of global modes in providing access to functional conformers. For example, upon deforming GluA3 NTD along ANM *mode 1* alone, the RMSD from the GluN1-N2B heterodimer decreases from 13.06Å to 6.12Å (more than 50%).

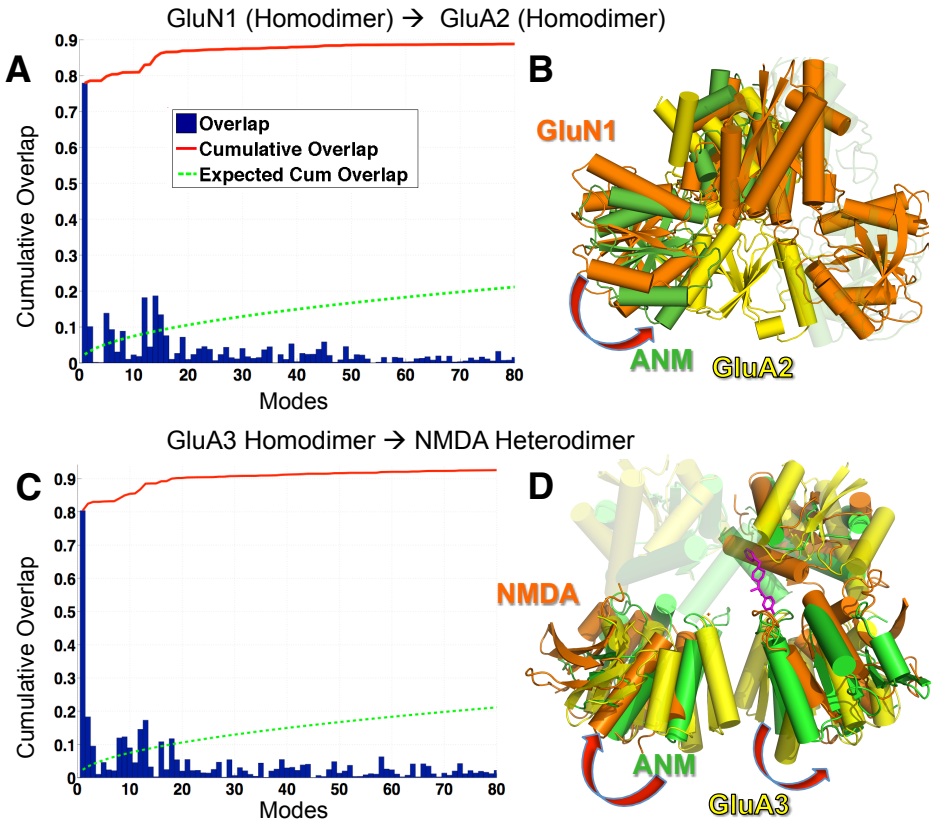


Figure 21. Ease of transition between dimeric conformers of NMDAR and AMPAR NTDs.

(A) Results are illustrated for the passage from GluN1 (NMDA) homodimeric conformer to GluA2 dimer conformer. The overlap (*blue* bars) represents the correlation cosine (see *Methods*) for each of the top-ranking 80 ANM modes to the conformational change. The *red* curve represents the cumulative overlap, adding up the contribution of all modes starting from the low frequency end (*mode 1*). The dashed green curve displays the *control*, for random modes. The slowest mode predicted for GluN1 (PDB 3Q41) yields an overlap of ~80%, indicating a strong pre-disposition of the GluN1 homodimer to assume the conformation of the GluA2 dimer. **(B)** Two transitional end points (*orange*, *yellow*) and an intermediate structure reached by moving exclusively along *mode 1* (*green*). **(C, D)** Same as A and B, for the change in the conformation of GluA3 homodimer (*yellow*) toward that of the heterodimer GluN1/GluN2B (PDB 3QEL, *orange*) along GluA3 ANM intermediate (*green*).

Results for other pairs of conformers between AMPAR and NMDA subfamilies are shown in **Figure 22.** 90% cumulative overlap with the targeted deformation (*red* curve) is attainable with a small subset (e.g. 20-25) of soft modes (out of a total of ~1,800 ANM modes) in all cases, except for the GluA2 → GluN1 homodimer, in which the overlap is ~ 70%.

Together with the ANM data (**Figures 18-21)**, these results underscore an unexpected parallel between AMPAR and NMDAR NTD flexibility and dynamics.

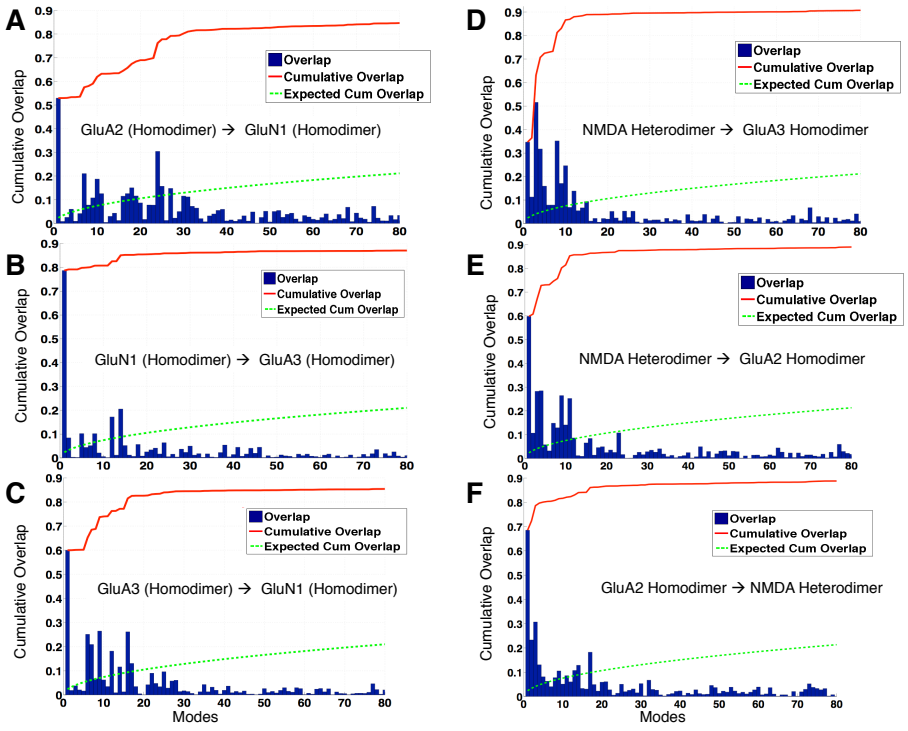


Figure 22. Ease of transitions between dimeric conformers of NMDA and AMPA receptors NTDs. (A) The back transition from GluA2 to GluN1 is shown here and can also be explained by ANM modes (90% using 80 modes). (B and C) show the transition from dimeric GluN1 to GluA3 and back from GluA3. (D) Shows transition from NMDA heterodimer to GluA3, the top 80 modes of the NMDA receptor can achieve 90% of the deformation (E) Shows the transition from NMDA heterodimer to GluA2 and (F) from GluA2 to NMDA heterodimer.

3.5 GLUA3 CAN ALSO TRANSIT INTO MGLUR1 CONFORMATIONS

Using a similar premise as in Section 3.4.3, we examined if the structural differences between GluA2/GluA3 and the mGluR1 *apo* form (PDB 1EWT) are conformations that are 'accessible'

via modes intrinsically favored by these structures. The *apo* form of mGluR1 features a large (~70°) rotation about the dimeric interface and is the mGluR structure structurally most different to the iGluR NTDs (RMSD of ~ 14Å); furthermore, the *apo* form exhibits the functionally relevant dimeric rearrangement, which we wished to test for accessibility to GluA3. The passage from GluA3 to mGluR1 conformations, can also be achieved upon displacement along a small subset of soft GluA3 modes (**Figure 23A-B**), although the contribution of the softest mode (mode 1) in this case is slightly smaller (~0.55). The cumulative overlap (or correlation cosine) between mode 1 of mGluR1 and GluA3 is 0.83. This suggests that there is a direct path connecting even the most extreme conformation of mGluR1 to GluA3 in conformational space (**Figure 23C-D**). Interestingly, this path is accessible via mGluR1's most dominant modes, allowing a reconfiguration of mGluR1 into GluA3 and matching the allosteric rearrangement of the mGluR LBC upon glutamate-binding.

The results obtained for all transitions between the three structures indicate that the experimentally resolved GluA2/3 and mGluR1 structures essentially represent conformers in a subspace most readily accessed via their soft modes, which are naturally favored by the shared overall architecture between these receptors. We note that the passage between GluA2 and GluA3 structures requires higher modes, as the structural difference between these two conformers is relatively small and involves more localized (as opposed to global) changes (**Figure 23A**). In summary, the analysis underscore intrinsic similarities between the dynamics of AMPAR NTDs and the mGluR LBC (as seen for NMDARs); the closer 'functional' relationship between mGluR1 and GluA3 can be explained by their similar LL arrangements.

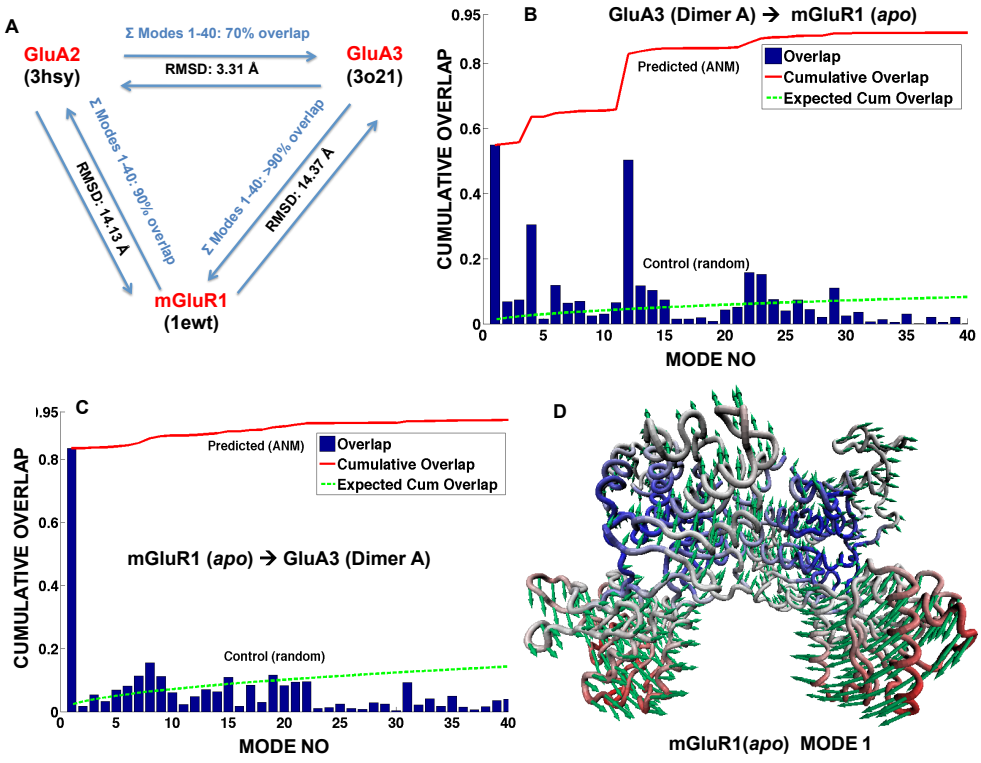


Figure 23. Transitions between AMPAR and mGluR1 conformations.

(A) The conformations of GluA2, GluA3, and mGluR1 are all mutually accessible. The normal modes of the three proteins show that all three form a subspace within the conformational landscape and that transitions are possible between the three, facilitated by the softest modes. The first 40 modes show significant cumulative overlap between the iGluRs and the mGluR (percentage of cumulative overlap is indicated for each transition), indicating that global, collective dynamics are sufficient to allow iGluRs to access mGluR mobilities. Note that GluA2 and GluA3 show less overlap, indicating that higher modes (i.e. more localized deformations) are necessary for the transition **(B)** Shows explicitly the overlap for transition from GluA3 (40 of 1953 modes shown) to the mGluR1 (*apo*) form. The deformation vector was calculated between 653 aligned residues. Mode 1 shows the highest correlation cosine with an overlap of 0.55, and an overlap of >90% is achieved by the first 40 modes (red line). The green line shows the expected cumulative overlap in case of random modes, indicating how well predicted modes do as compared to the control and that there exists a noticeable path of transition from GluA3 to the mGluR1 (even in its highly deformed *apo* form) via the slow modes. **(C)** Shows overlap for transition from metabotropic receptor in the *apo* form to GluA3 (40 of 1953 modes shown) The deformation vector was calculated between 653 aligned residues. Mode 1 shows the highest correlation cosine with an overlap of 0.83, and an overlap of >90% is achieved by the first 40 modes (red line). The green line shows the expected cumulative overlap in case of random modes, indicating how well predicted modes do as compared to the control. **(D)** Motion of mGluR1 in mode 1 of ANM. The arrows clearly indicate the opening and closing of the lower lobe of receptor, with some smaller motions in the upper lobe. This opening and closing which is easily accessible to mGluR1, provides a transition path to its GluA3 like conformation.

3.6 LOCAL RESIDUE DYNAMICS FROM MD SIMULATIONS

3.6.1 High intra- and inter LL mobilities in AMPAR NTDs from MD simulations

To obtain a better understanding of the molecular interactions that underlie iGluR NTD dynamics, we conducted all-atom MD simulations. Root-mean-square fluctuations (RMSFs) in residue positions (**Figure 24**) confirm that GluA3 exhibits the highest mobility among all AMPAR NTDs. This enhanced mobility is primarily mediated by LL helices αE and αF, in agreement with data from fluorescence correlation spectroscopy experiments (Jensen *et al.*, 2011). These helices may make contacts with the LBD in the intact structure of AMPAR (**Figure 15A**) and could thus communicate to downstream segments of the receptor. Similarly, helix αH located next to the entrance of the cleft in AMPAR NTDs features high mobility, consistent with the structural variation observed upon superposition of GluA1-4 structures (**Figure 14A**).

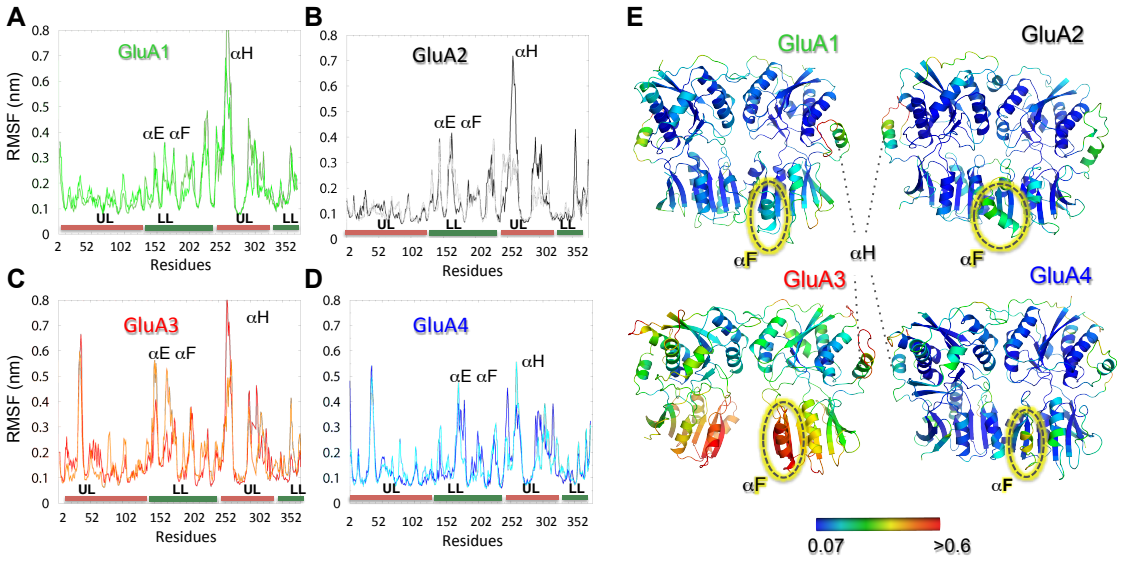


Figure 24. Fluctuation profiles of residues for AMPAR NTDs.

(A-D) RMSFs plotted as a function of residue index. The helices with high mobilities (αE, αF and αH) are labelled. We also note the suppressed motions of GluA1 and GluA2 at the UL N-terminal portions. GluA3 and GluA4 exhibit

peaks near A36-T37 (on the loop between αA and β2; not labeled). The bars below the curves in **A-D** indicate the UL (*red*) and LL (*green*) segments along the sequence. In panel E, the respective structures are color-coded according to their residue RMSF values plotted in panels A-D, from least mobile (*blue*) to most mobile (*red*). The most mobile helix, αH, is indicated. The helix αF is encircled to highlight its high mobility.

Next, we monitored interlobe (UL-UL and LL-LL) distances based on their centers of mass (CMs). Simulations clearly show that the UL-UL distances (~ 4.0 - 4.3 nm) are maintained in all AMPAR NTD dimers (inset of **Figure 25A**), while LL-LL distances vary between dimers: they maintain their original values (of 2.9 - 3.3 nm) in GluA1, 2 and 4. In GluA3 however, they increase to more than 4.5 nm at early stages of the simulation, essentially disrupting the LL interface. Snapshots of the GluA3 NTD at different stages (**Figure 25B**) illustrate the loss of the LL interface within the first 5 ns, followed by stabilization of a different conformation

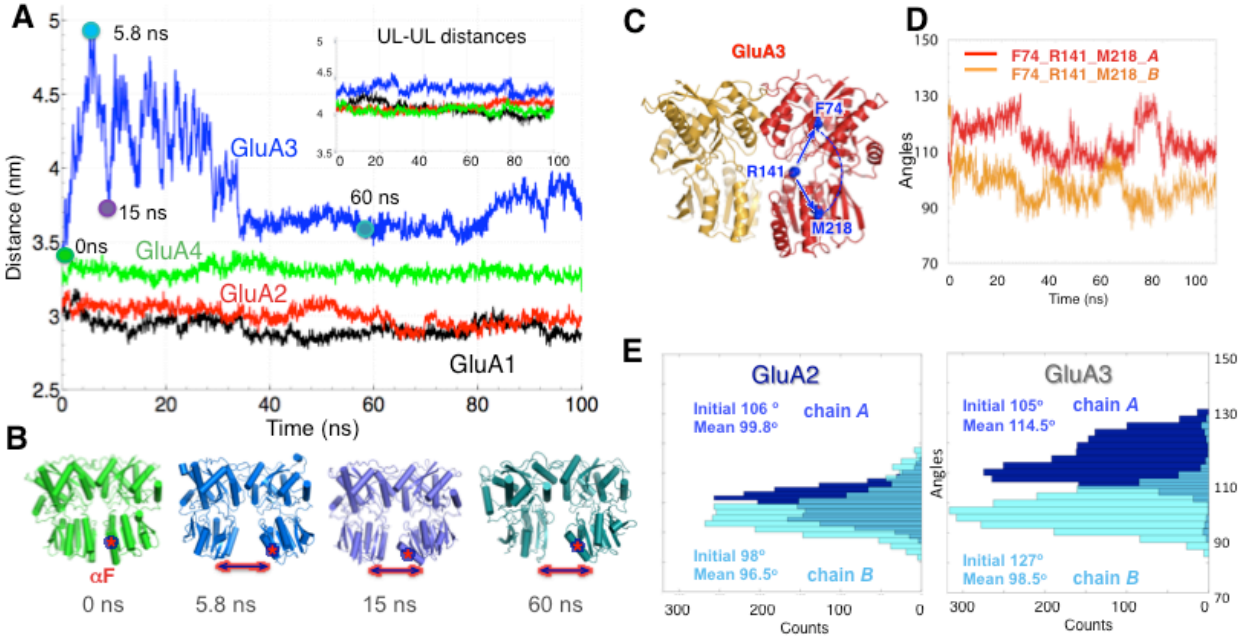


Figure 25. Lower lobe interface instability of GluA3 evidenced by comparative analysis of MD simulations for GluA1-GluA4.

(A) Distance between the mass centers of LLs, shown for GluA1-GluA4 NTDs as a function of simulation time. Results for the ULs are shown in the inset. Large fluctuations are observed in GluA3 LL-LL distance (*blue* trace).

(B) Snapshots display GluA3 conformations at t = 0, 5.8, 15 and 60 ns (see colored circles in A). **(C)** Probe residues selected for monitoring the changes in interlobe cleft angle, shown for GluA3 NTD (3O21). **(D)** Time evolution of inter-lobe angle for GluA3 protomers. Note the periodic opening/closing and the anticorrelation between the protomers. **(E)** These properties are contrasted to those observed for GluA2 where the angles exhibit minimal fluctuations. Histograms refer to inter-lobe angles for protomers ***A*** (*dark blue*) and ***B*** (*cyan*).

distinguished by the loss of αE helicity and the reorientation of αF towards the LBD. As shown below, this behaviour is due to the unique positive charge distribution in the GluA3 LL interface (Sukumaran *et al.*, 2011). GluA3 also exhibits localized rearrangements in the UL dimer interface, which are not seen in the other AMPAR paralogs. Specifically, hydrophobic packing is disrupted as phenylalanine pairs (F56, F88) are separated and in some cases, irreversibly broken (**Figure 26)**. This "acquired" UL instability further points to the unique behaviour of the GluA3 NTD, potential coupling between the LL and UL interfaces in AMPAR NTDs, and the importance of the LL as a key structural determinant mediating intrinsic dynamics.

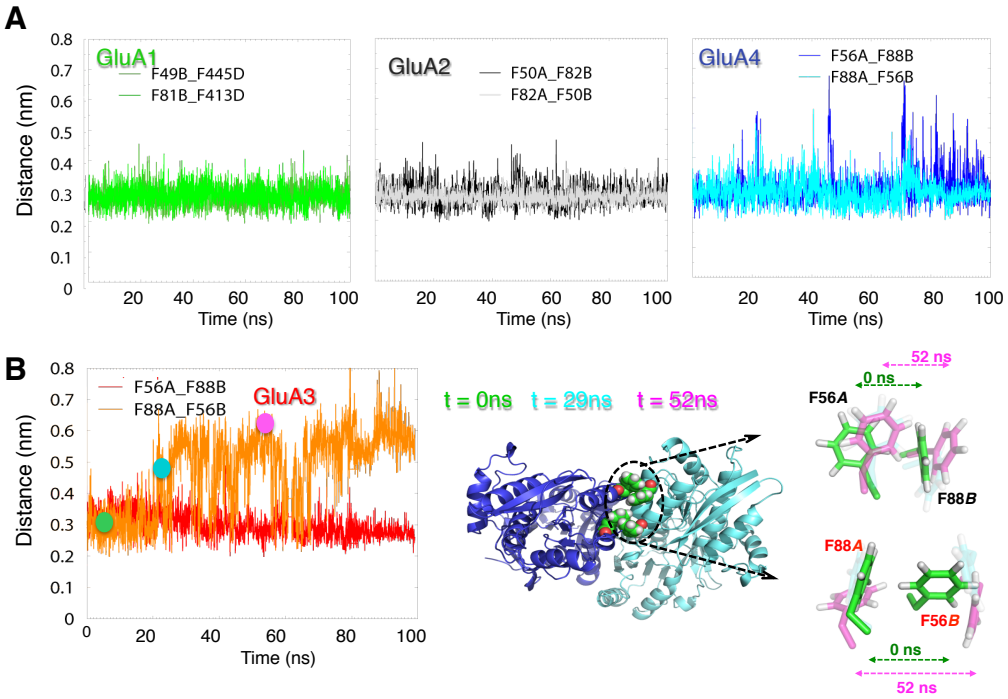


Figure 26. UL interface stabilities for GluA1-4

(A) We show the distances measured between Phe residues at the UL interfaces, between the two protomers, in each of GluA1, GluA2 and GluA4. There are two pairs of the Phe-Phe interaction at the UL interface, which are both maintained in GluA1, GluA2 and GluA4. (B) The same distance is shown for GluA3. In GluA3, one of the interaction weakens at ~30ns, after which it exhibits large fluctuations. The snapshots of the interacting pairs of Phe (F56*A*_F88*B*, and F88*A*_F56*B*) in GluA3 are illustrated at t=0ns, 29ns and 52ns. While the F56*A*_F88*B* interaction is maintained, the F88*A*_F56*B* interaction is disrupted, mainly due to movement of F56*B*.

3.6.2 GluA3 NTD protomers undergo clamshell-like motions

To determine the influence of interface stability on classic PBP-like clamshell motions (Quiocho *et al.*, 1996;Trakhanov *et al.*, 2005) we examined the fluctuations in the inter-lobe cleft angle, based on three C^{α}-atoms in each AMPAR NTD (**Figure 25C**). We observe markedly larger angular fluctuations in GluA3 than in GluA2 (**Figure 25D-E**); GluA3 featured the widest opening of inter-lobe cleft angle (ranging up to > 130^{o} in protomer *A*). Interestingly, the two GluA3 protomers appear to undergo anticorrelated fluctuations, with protomer *A* closing and *B* opening with a periodicity of ~25 ns (**Figure 25D**). This motion is unique to GluA3 and not discernable in other AMPAR counterparts; suggesting that LL flexibility in GluA3 may be coupled to clamshell-like motions of the individual protomers. Moreover, cleft motions in GluA3 (**Figure 25D**, orange curve) are accompanied by changes in UL hydrophobic packing (**Figure 26**), together suggesting a coupling between clamshell-like motions of the individual protomers and inter-protomer packing.

3.6.3 Effect of LL residues on interlobe packing and dynamics

The difference in GluA2 and GluA3 interface stability and residue fluctuations observed in MD simulations is also reflected in their dimer stabilities derived experimentally (Rossmann *et al.*,

2011). In GluA2, hydrophobic residues contribute to the LL-LL contacts while in GluA3 pairs of arginines (R163 and R184) project into the interface, destabilizing the dimer (Sukumaran *et al.*, 2011). To gain further insight into the relationship between the interface stability and cleft dynamics, we analyzed two mutants generated *in silico*: L144D (GluA2), to destabilize the GluA2 interface via like-charge repulsion; and R163I (GluA3) to strengthen the labile GluA3 interface via hydrophobic contacts.

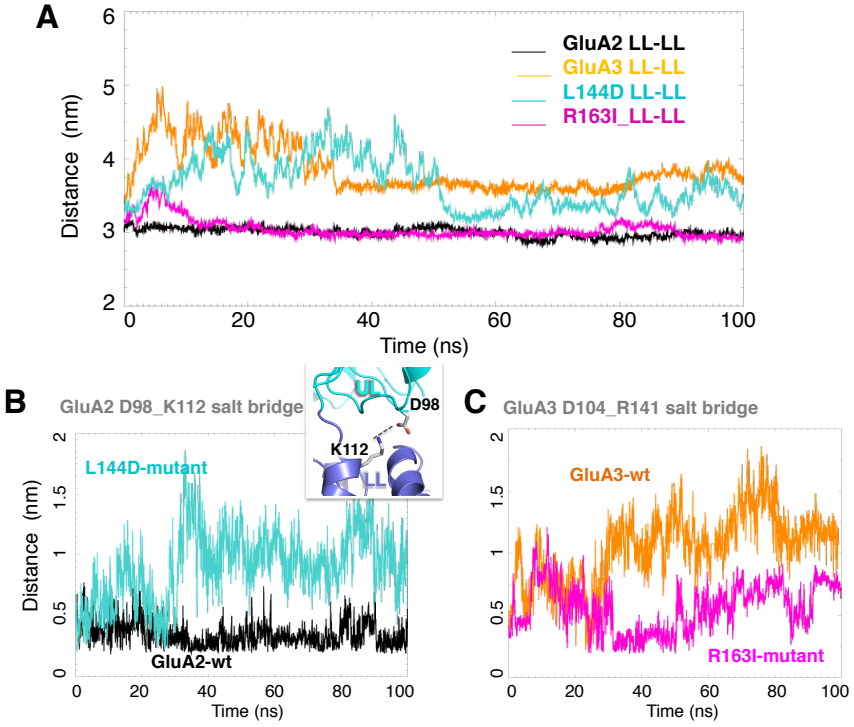


Figure 27. Critical role of inter-residue interactions at LL-LL interface in defining dimer dynamics.

Results are presented for the mutants L144D (GluA2) and R163I (GluA3) to examine the significance of hydrophobic *vs* charged interactions in defining the distinctive dynamics of GluA3 and GluA2. (A) Time evolution of the closest inter-atomic distance between L144 residues on neighboring subunits for the wildtype (*black*) and between D144 pairs in the mutant (*teal*). Inset highlights the region of mutation. (B) Snapshots of wildtype GluA2 and L144D mutant at 100ns, superimposed and viewed from bottom. (C-D) Same as (A-B) for GluA3 wildtype and mutant R163I.

Introduction of like-charges into the GluA2 LL interface indeed led to destabilization as can be seen from the comparison of the LL-LL distances for the mutant L144D (*teal* curve in **Figure 27A)** and for the wildtype GluA2 (*black* curve). Notably, the extent of destabilization is comparable to that originally observed for GluA3: inter-lobe distance between the two substituted amino acids increases to more than 30Å within tens of nanoseconds; whereas in wt GluA2, the equivalent interaction is maintained over a period of 100 ns. Conversely, introducing hydrophobic residues into the GluA3 LLs leads to a more stabilized interface (**Figure 27C**): the distance between the mutated R163I residues is maintained, but is disrupted in wt GluA3 at early stages of the simulation. Snapshot at t = 100 ns illustrates the disruption of the LL packing interface upon L144D mutation in GluA2 (**Figure 27B**), and strengthening in R163I (**Figure 27D**). These results thus demonstrate the stabilizing role of hydrophobic residues at the packing interface in GluA2, as well as the destabilizing role of buried arginines in GluA3. CM distances between the LLs (**Figure 28A**) further establish that the GluA2-L144D mutant weakens dimer contacts, while GluA3-R163I is stabilizing.

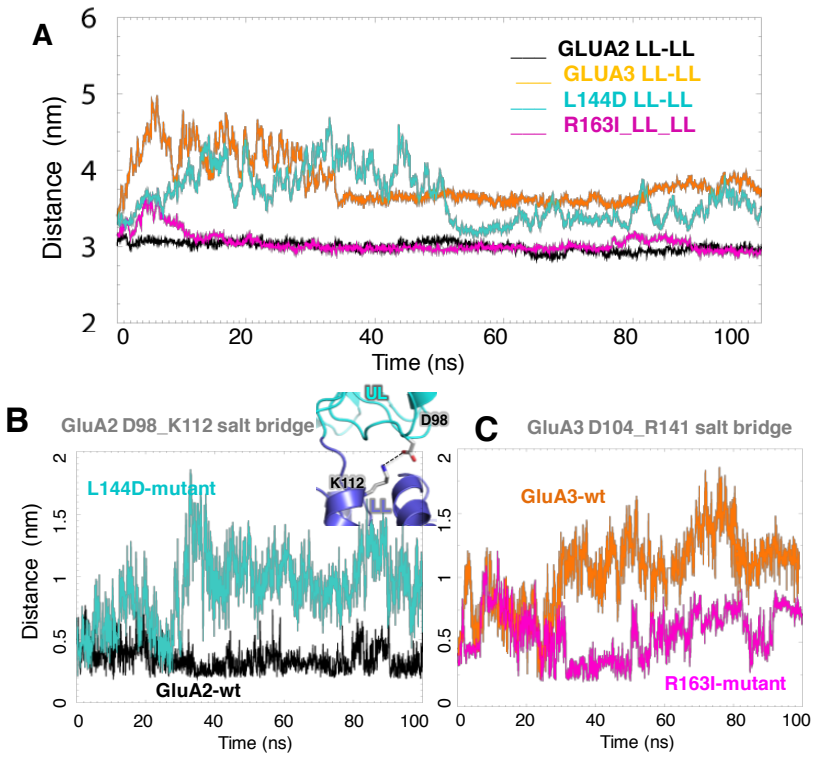


Figure 28. Time evolution of interlobe distance observed for wild type and mutant GluA2 and GluA3.

(A) Distance between the LL mass centers for GluA2 (*black*), GluA3 (*orange*), GluA2 mutant (L144D) (*teal*) and GluA3 mutant (R163I) (*magenta*). GluA2 WT and GluA3 mutant R163I are highly stable, indicated by the constant distance maintained, at ~ 3nm. On the other hand, the destabilizations of the LL interface in GluA3 WT and GluA2 mutant L144D are evident during the early stages of the simulation, lasting up to ~ 40ns. **(B)** Distance between UL-LL salt-bridge forming residue bresidues pairs D98 and K112 for GluA2 (*black*) and GluA2 L144D mutant (*teal*). The distance is measured in terms of the minimum distance between any two atoms of the two residues. **(C)** same as **(B)** between UL D104 and LL R141 for GluA3 (*orange*) and GluA3 R163I mutant (*magenta*).

Analyses of the trajectories generated for the mutants show that LL stability is coupled to UL-LL dynamics: a salt bridge connecting the lobes of the clamshell (D98-K112; **Figure 28B**, expected to restrain cleft motions, is destabilized in the GluA2 mutant, whereas the R163I mutation stabilizes the equivalent salt bridge in GluA3 (D104-R141; **Figure 28C**). Moreover, stabilization of the LL in GluA3-R163I restricts clamshell motions as compared to the wt GluA3 dimer. Therefore, alterations of LL interface strength can propagate to the hinge region in both

cases and has the capacity to alter lobe motions; i.e. the perturbation of LL stability in both GluA2 and GluA3 has bidirectional effects that extend beyond local (LL) interactions.

3.7 DISCUSSION

In this chapter, we provide a series of novel insights into the dynamics of AMPA- and NMDA receptor NTDs. First, we present the crystal structure of GluA3 and GluA4 NTD, facilitating a comprehensive analysis of this sequence-diverse domain across the AMPAR subfamily. Secondly, we provide mechanistic insights into the intrinsic dynamics of GluN2B that facilitate ligand binding and offer a first glance into the motions driving GluN2B NTD allostery, whose *modus operandi* has not been elucidated to date. Third, we reveal that AMPAR- and NMDAR NTD monomers share surprisingly similar global mode motions. These are restricted, but not abolished, upon dimerization in a subunit-dependent fashion, dictated by the evolutionary and structurally variable LL interface. Fourth, we show that AMPAR NTDs can readily re-configure into NMDAR NTD as well as mGluR LBC conformers, a further indication of their unexpected similarity and their putative allosteric capacity. Fifth we evaluate the dynamics of the AMPAR subfamily at atomic resolution, where, in accordance with experimental data, we find that GluA3 features the weakest LL dimer interface, which ruptures after ~ 5 ns of MD simulations followed by GluA3 LL secondary structure elements (αF) flipping downward to the LBD. The unique LL packing of GluA3 also potentiates it to undergo classic PBP-like clamshell motions. Finally, we capture critical residues at the LL-LL interface that mediate inter-protomer dynamics in AMPARs, consolidated by analyses of mutants designed to weaken or strengthen the LL interface.

The AMPAR NTDs studied here are stable homodimers in solution, with a highly conserved UL interface, which will maintain dimer stability. The LL, on the other hand, which potentially shares an interface with the LBD, may play a mediatory role in the allosteric regulation, also demonstrated by the recent NMDAR NTD heterodimeric structure (Karakas *et al.*, 2011). In AMPARs, helices αE and αF along with β7 together form the LL dimer interface in most crystal structures. Previous work has established that the GluA3 NTD assembles into homodimers less tightly and preferentialy co-assembles with other AMPAR NTDs into heterodimers; the weak homomeric LL interface underlies the distinctive dynamics of the GluA3 NTD (Rossmann *et al.*, 2011;Sukumaran *et al.*, 2011). Indeed, in our simulations of all AMPAR NTD homodimers, GluA3 is distinguished by its high mobility: αE (L137-K151) shows considerable unwinding; αF tilts towards the NTD/LBD interface. A partial loss of helicity in αE is also observed in the recent structure of a kainate receptor (GluK3) NTD (Kumar and Mayer, 2010), which also assembles as obligate heteromers, supporting the link between enhanced mobility (or lower stability) at the LL-LL interface and low homodimeric assembly propensity. The instability of the GluA3 LL dimer interface may propel towards the UL interface, which is apparent in MD simulations (**Figure 26**). The observed loosening of the hydrophobic core (F56 and F88) may facilitate interprotomer rotations. The downward motion of helix αF towards the LBD in MD trajectories suggests a potential role in the allosteric propagation of NTD motions. The cross-talk between NTD and LBD will be affected by the connecting linker. This segment is sequence variable between the paralogs and harbors two N-glycosylation sites. These have been removed in the GluA2 homomeric structure (PDB 3KG2) along with a deletion encompassing six residues (Sobolevsky *et al.,* 2009). How this mutation affects domain packing and allosteric communication in iGluRs is a key open question.

Our analysis reveals the ability of individual protomers to undergo concerted clamshell opening/closing motions, which simultaneously affect inter-protomer contacts. This supports a possible cooperative response of AMPAR NTDs upon ligand binding or interaction with protein partners (O'Brien *et al.*, 1999) and a capacity to transmit signals towards the channel. We note that ligand interaction may not be restricted to the interlobe cleft but could target the LL dimer interface, as known from analogous cases (He *et al.*, 2006; Mony *et al.*, 2011), or the highly dynamic αH region.

The NTD of the NMDAR is known to modulate channel gating by binding Zn^{2+} ions and ifenprodil-like compounds, thereby sparking clinical interest in these domains. The twisted LL along with surface properties (Karakas *et al.*, 2009; Stroebel *et al.*, 2011) have been purported to be the primary reason why NTD-mediated modulation of the ion channel is seen in GluN2B and NMDAR heterodimers, but have not been seen so far in non-NMDA receptors. Structural dynamics analysis offers a different perspective, where the global motions accessible to the different NTD structures of the iGluR families overlap remarkably. Also, the global modes of dimeric AMPAR, NMDAR, and mGluR NTDs allow for facile transitions from one form to another, suggesting that the AMPAR NTDs may equally have allosteric signaling abilities.

Binding of ions and small molecules to the NTD are most likely facilitated by global motions in NMDARs. NTD clamshell motions have been implicated in facilitating an induced-fit binding mechanism (Karakas *et al.*, 2011). Based on similarity of global motions between AMPARs and NMDARs, the allosteric effect known to modulate NMDAR open probability should not be disregarded for non-NMDARs. This view is further strengthened by the small-molecule binding capacity in the GluA2 cleft reported previously (Sukumaran *et al.*, 2011), and the labile nature of the GluA3 NTDs observed here. The present analysis is a further step toward

clarifying the putative allosteric potential of AMPAR NTDs by highlighting their intrinsic ability to undergo motions comparable to NMDAR NTDs, and their propensity to sample conformers observed in NMDARs. This opens an avenue of searching for molecules able to bind AMPAR NTDs, which may in turn play an important role in regulating gating of nonNMDA iGluRs.

This study was performed jointly with our collaborator Dr. Ingo Greger and with Dr. Indira Shrivastava that has the results have been published in Dutta et al., 2012 and Sukumaran et al. 2011. Having characterized the dynnamic potential of AMPAR NTDs, we seek to understand the behavior of individual domains in the context of the whole receptor. We also proceed towards the identification of a potential NTD mediated allosteric communication pathway in AMPARs.

4.0 A SERIES OF RESIDUES ACTING AS SENSORS AND EFFECTORS REGULATE ALLOSTERIC COMMUNICATION IN THE INTACT AMPA RECEPTOR

All pharmacologically distinct classes of iGluRs include N-methyl-D-aspartic acid harbor an extracellular domain (ECD), a transmembrane domain (TMD or ion channel) and an intracellular carboxyl-terminal domain (CTD). The ECD of iGluRs consists, in turn, of a ligand-binding domain (LBD), and an N-terminal domain (NTD) (Traynelis et al., 2010) (**Figure 29A**).

A large body of work on the LBD, owing to availability of high quality structural data, suggests that iGluRs receptor gating is tightly coupled with the dynamics of the LBD. These studies indicate that the clamshell-like closure of the bilobate structure of LBD monomer (S1-S2) upon glutamate (or agonist) binding allosterically triggers the opening of the TMD pore (Erreger et al., 2004; Mayer, 2006). The degree of domain closure, and accompanying extent of channel activation is ligand-dependent (Armstrong and Gouaux, 2000; Jin et al., 2003; Jin and Gouaux, 2003). Receptor deactivation takes places by two possible mechanisms: ligand unbinding, which induces closure of the TM channel, or conformational rearrangement of the two monomers of LBD dimer with respect to each other which leads to the de-sensitized state (Armstrong et al., 2006; Hansen et al., 2007; Horning and Mayer, 2004; Sun et al., 2002). However, the exact mechanism of pore opening coupled to LBD lobe closure remains to be established.

Like the LBD, the NTDs are bi-lobal and they form stable dimers. The compact structure of AMPAR NTD lower lobes (LL) suggests AMPAR NTDs to be rigid entities without a regulatory role in ion channel gating (Jin et al., 2009; Kumar et al., 2009). In contrast, the LLs of NMDAR heterodimer NTDs hardly make interlobe contacts. Their loose packing enables clamshell-like motions in the NMDAR subunit GluN2B, along with large-scale twisting movements in the subunit NR1 (Zhu et al., 2013), which presumably facilitates ligand/ion binding. This structural flexibility imparts the NMDAR NTD with channel modulation properties (Gielen et al., 2009; Karakas et al., 2009; Karakas et al., 2011; Sirrieh et al., 2013; Zhu et al., 2013). However, two recently resolved crystal structures of AMPAR NTDs, GluA4 and GluA3 permitted us to comparatively examine the NTD conformational dynamics of all subtypes (GluA1-4) of AMPARs (Chapter 3). The study revealed that AMPAR NTDs, in contrast to early assessments, do possess intrinsic conformational flexibilities that favor clamshell-like motions as seen in NMDAR NTDs and homologous LBD of mGluRs. This hints at a putative allosteric potential in AMPAR, akin to NMDAR NTDs.

The crystal structure of the GluA2 AMPA receptor (AMPAR) resolved so far (NTD, LBD and TMD), provides an opportunity to gain insights into the functional properties of the intact receptor, which otherwise has been limited to the individual domains in isolation (Sobolevsky et al., 2009). This structure revealed, for the first time, the modular architecture and atomic structure of the intact tetrameric receptor (except the CTD) in the closed form (of the TMD) (**Figure 29A**). Information on the intact structure is important because the dynamic behavior of the individual domains in the intact structure may differ from that in the isolated form. This shows that iGluRs are dimers of dimers where the ECD has an overall two-fold axis of symmetry, while the TMD exhibits four-fold symmetry. This hallmark structure offers a

unique platform to analyze the collective dynamics of all three domains crystallographically resolved (NTD, LBD and TMD), and elucidate the interdomain and intersubunit couplings that mediate AMPAR function.

In this chapter, we provide a comprehensive assessment of the intrinsic dynamics of the intact AMPAR. We focus in particular on a global mode of motion that induces correlated movements between the NTD and the LBD, while the LBD undergoes twisting motions in concert with the TMD. This type of coupling across the entire structure, predicted by the anisotropic network model (ANM) based normal mode analysis (NMA) (Atilgan et al., 2001; Bahar et al., 2010b), may be instrumental in propagating allosteric signals all the way from the NTD to the TMD. Next, we examined if/how the intrinsic dynamics of the individual domains are maintained in the intact structure. We found that this was the case for the NTD dimer, but not, for the LBD dimer. In particular the clamshell-like motions that require coupled opening and closing of the respective LBD and TMD cannot be sampled in the intact structure, due to locking of the antagonist-bound LBD in an open conformer. We further examined, with the help of residue-perturbation-scanning (PRS) analysis (Atilgan and Atilgan, 2009), the key sites that potentially serve as sensors and effectors for receiving and transmitting signals. Drawing additional information from graph theoretic methods, we identify a sequence of residues that define an allosteric communication pathway between the NTD and the TMD gate, which opens new avenues for modulating iGluRs' function.

4.1 INSIGHTS FROM ANM ANALYSIS OF THE INTACT AMPAR

4.1.1 Collective motions that couple the three domains of AMPAR

Primary highlights from the intact crystal structure include (i) the crossover of subunits from the NTD to the LBD (**Figure 29A**) such that monomers A-B and C-D that form coherent dimers in the NTD swap partners to dimerize as A-D and B-C in the LBD, and (ii) the symmetry mismatch between the NTD/LBD and the TMD (from 2- to 4-fold). These features yield two conformationally distinct subunit pairs: the distal pair A/C, and the proximal pair B/D. Notably, this difference in the quaternary packing of subunits is also reflected in their dynamics, as described below.

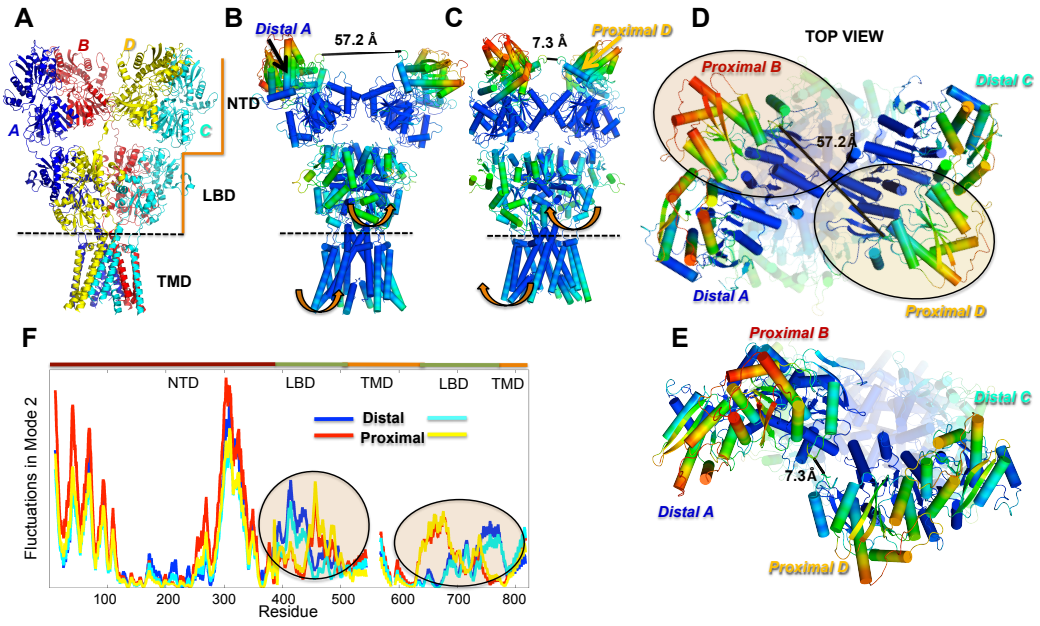


Figure 29. Intrinsic dynamics of the AMPAR captured with ANM global mode 2.

(A) Crystallographically resolved AMPAR (subtype GluA2) structure (PDB ID: 3KG2) (Sobolevsky et al., 2009) displaying the packing of chains A (*blue*), B (*red*), C (*cyan*) and D (*yellow*) in three domains, NTD, LBD and TMD; (B-C): Side view of two alternating conformers visited by AMPAR along ANM mode 2 (see Supplementary Material). Distance changes based on a representative residue (K268 in the αH helix on subunits B and D) highlights

how the proximal subunits of the NTD dimers undergo large distance fluctuations between 57.2 Å and 7.3 Å. This mode entails a global twisting in the LBD and TMD shown by the orange arrows. (D-E): Top view of the same conformers as in B-C, where the NTD is highlighted and the remaining receptor is in faded color, with the proximal subunits being highlighted by the ellipses. Each of the diagrams in B-E is colored by the mobility (normalized square displacement) of residues shown in panel F, from *red* (most mobile) to *blue* (most rigid). Ellipses show regions of the LBD where the distal and proximal chains exhibit distinctive dynamics.

The ANM analysis permits us to decompose the dynamics into a series of collective modes. Those at the lowest frequency end, also called *global modes*, usually describe the most cooperative movements encoded by the overall architecture. Application to AMPAR (subtype GluA2) revealed a potentially functional motion (mode 2; **Figure 29B-C)**, which encompasses all three domains, NTD, LBD and TMD. In this mode the pair of NTD dimers alternate between 'distant' (**Figure 29A** and **C,** *side view* and *top view, respectively*) and 'close' forms (**Figure 29B** and **29D**), while the LBD undergoes a twisting motion, coupled to the accompanying torsional rotation of the TMD (see *arrows* in **Figure 29B-E**). Note that the ANM provides robust information on the direction of motion, while the absolute size of the movement depends on the choice of the spring constants. We adopted here a spring constant compatible with the experimentally observed Debye-Waller factors in the crystal structure, which led to fluctuations of ± 25 Å in the indicated distance between the upper lobes of the NTDs.

In **Figure 29B-E**, the receptor is colored by the mobility profile (square fluctuations) of individual residues, displayed in panel **F**. Peaks in panel F (*red regions* in **Figure 29A-D**) show regions of high mobility; whereas minima (*blue regions* in **Figure 29A-D**) represent residues with inhibited motions. The mobility profiles of the distal and proximal subunits are similar at the NTD, but differ in the LBD (encircled regions), due to their particular packing characteristics. In particular, the subunit regions at the LBD tetramer interface are less mobile, whereas their symmetric counterparts exposed to the aqueous environment enjoy higher mobility

(**Figure 30A-B**). The twisting in the LBD is also propagated downward, such that there is a rigid-body twist in the TMD (**Figure 29A-B**, *arrows*). We note that the TMD motions may be more constrained than those predicted by ANM due to the presence of the lipid bilayer.

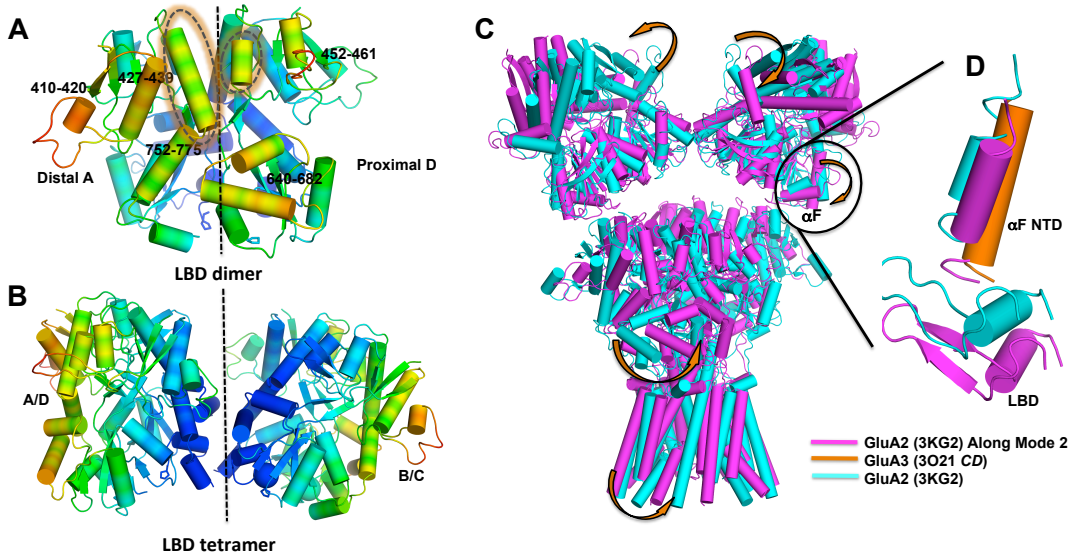


Figure 30. Characteristic motions in the NTD and LBD captured by slow mode 2 motions of ANM.

(A-B): The differences in the mobility of the LBD proximal and distal chains (highlighted in **Figure 29F**) are described here by color-coding the LBD with mobility in slow mode 2. In the LBD, residues of the helix 742-757 form interfacial contacts with residues in the helix–loop region of 482-495. A shows the front view of this interfacial region where the helix 742-757 in distal chain A (encircled) comes together with helix–loop region 482-495 of proximal chain D. Also, regions of the LBD distal and proximal chains that show high mobility are highlighted (numbered residues). **B** shows the side view of the whole receptor LBD, where we see that the A/D dimer subunit shares a tetramer interface with B/C subunit of the LBD. This tetramer interface between subunits results is dampening of the motions at this otherwise symmetric end of the LBD dimer (distal-proximal pair). Thus the symmetric intrasubunit distal-proximal interface formed now by helix 742-757 of proximal chain D and helix–loop region 482-495 of distal chain A, show lesser mobility than their respective counterparts shown in A. Hence regions numbered in A for distal chain (410-420, 427-439, 752-775) show dampened motion in the proximal counterpart of the symmetric interface (that facing the tetramer interface), and similarly for those shown in A for proximal chain (452-461, 640-682). This accounts for the difference in mobilities highlighted in **Figure 29F**. **(C)**: The structure of the intact receptor and one of the two end-point conformations obtained by deforming the receptor (PDB ID: 3KG2, along ANM mode 2 (**Figure 29B**, shown here in *magenta*) are superimposed on top of each other. The arrows indicate the direction of motion, as described in **Figure 29**. The circled region shows the motion of the αF helix, that moves downward as the proximal NTDs move far away (*magenta*). This is highlighted further in **(B)**. Note that, we

see a downward motion in the proximal αF helix, when the NTDs come "close" to each other. However, here we focus on the distal αF helix, since the downward motion brings it closer to the LBD (not seen in the proximal subunits due to domain swapping and a larger proximal NTD-LBD distance). **(D)** On superimposing the GluA3 αF helix (*orange*) on the crystal structure of the AMPAR (3KG2) (αF helix 172-189, distal chain *A*), we see that GluA3 αF shows a similar downward motion (as seen in the ANM deformed structure, *magenta*) (3O21-D 169-186), facilitated by a more open LL interface. Intact AMPAR global movements conform to structural change observed between GluA2 and GluA3 NTDs.

Figures 30C and **D** show that the structural change induced in GluA2 NTD along mode 2 conforms to the structural difference between the GluA2 and GluA3 subtypes of AMPAR NTD. Basically, the **Figure 30D** displays the NTD structure resolved for another AMPAR subtype, GluA3 (*orange*), superimposed on GluA2 crystal structure and ANM-predicted conformer. This comparison shows that the ANM-predicted downward motion of helix αF coincides with the conformational difference between GluA2 and GluA3. The downward motion in GluA3 is facilitated by a relatively open LL interface. The conformational flexibility of the αF helix in GluA3 has been previously noted (Dutta et al., 2012; Jensen et al., 2011) and is also mirrored in the GluA2 full-length AMPAR in a more global mode, thereby hinting at its role in allosteric communication.

4.1.2 NTD and LBD dynamics within intact receptor framework

In chapter 3 we have shown that the NTD of AMPARs can access two types of collective motions: clamshell motions of individual monomers, and counter-rotation of monomers within the dimer enabling access to NMDAR-like configuration. We probed the extent to which the motions seen in the isolated NTD dimers are preserved in the intact structure, with the help of a perturbation method (Ming and Wall, 2005). The method divides the examined structure into a

'system' and its 'environment' and permits us to compare the dynamics of the system in the presence and absence of environment (see *Methods*). Application to AMPAR NTDs yielded the results shown in **Figure 31**.

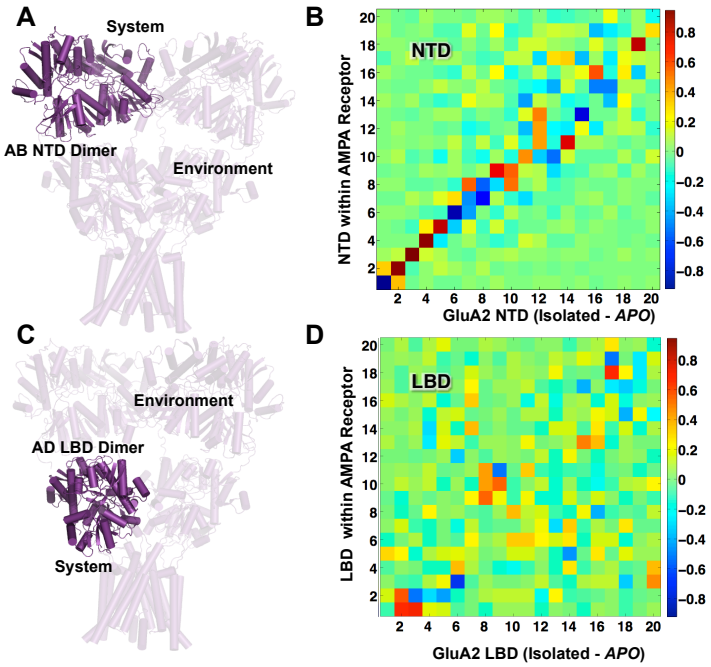


Figure 31. Comparing the dynamics of the isolated NTD and LBD within the framework of the receptor.

(A): This highlights the region of the receptor (AB dimer of the NTD) that is considered as the *system* (*deep purple*) and the *environment* (*faded purple*) for a perturbation analysis. (B): Correlation between top 20 modes accessible to GluA2 NTD in isolation (3HSY; *abscissa*) (Sukumaran et al., 2011) and in the receptor (3KG2: *AB* 10-384; *ordinate*). Darkest *red* and *blue* regions refer to strongest correlations, whereas *green* shows no correlations (see scale). (C) and (D) same as panels A and B, but for the LBD, where we probe correlation (D) between the dynamics of the isolated LBD *apo* dimer (1FTO; abscissa) with that of the LBD in the receptor (3KG2: *AD* 394-506, 632-773; *ordinate*).

First, we consider the AB dimer of the NTD as the system (**Figure 31A**, *purple*) and the rest of the receptor as its environment (**Figure 31A**, *faded purple*). **Figure 31B** shows the correlation map between the top 20 modes from the system in the presence of the environment (*ordinate*) and from the isolated GluA2 NTD (*abscissa*). Strikingly, first 10 modes of the isolated

GluA2 are almost identically preserved in the context of the overall structure (diagonal elements, with > 0.8 correlation cosine). Although the order of the higher modes is changed, we can identify counterparts that exhibit high correlations (departures from diagonal elements). Overall, these observations indicate that the NTD maintains its dynamic properties in the intact receptor.

Second, we examine the LBD dimer (**Figure 31C-D**). In the LBD, the degree of cleft closure of the bi-lobed LBD domain triggered by ligand binding is a key determinant of the potency of the ligand in inducing channel opening; agonists trigger large domain closure (and lead to open channel); whereas antagonists lock the cleft in a more open state (and in turn, stabilize the closed form of the channel, as seen in the current intact structure). **Figure 32** shows that in the absence of the environment, the domain closure of the *apo* (isolated) LBD is facilitated by ANM mode 1 accessible to the isolated LBD that enables access to a closed *holo-like* conformation. Comparison of normal modes predicted for isolated LBD (*apo* form) with those in the presence of the environment (AD dimer: **Figure 31C**) shows the lack of correlation (**Figure 31D)**. Thus, in contrast to the NTD, the motions accessible to the LBD dimer alone cannot be deployed in the intact receptor. Moreover, a few modes which do overlap with a correlation of ~0.6 are due to motions of the flexible loop arm (residues 409-421) and do not embody cooperative events.

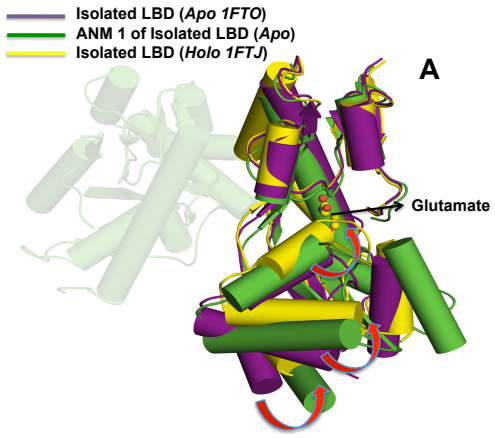


Figure 32. Transition from isolated *apo* form of the LBD to isolated *holo* form facilitated by ANM Mode1.

(A). This shows how isolated LBD *apo* structure (1FTO), when deformed along its ANM mode 1 (*green* intermediate configuration) is able to access a more closed form as seen in the *holo* glutamate-bound conformation of the isolated LBD (1FTJ). ANM modes have been shown previously to enable sampling of conformations that allow functional motions like those associated with ligand binding (Bakan and Bahar, 2009). The arrows indicate the direction of motion of some of the helices as they go from open to a more close conformation.

Overall, we see larger fluctuations in the NTD than the LBD in the global modes. This may be attributed to the loose packing of the NTD-LBD interface that enables the NTD to maintain its global dynamics within the receptor. In contrast, LBD motions in the currently resolved intact receptor are inhibited. The tight packing of the LBD/TMD linker helices and an overtly open form of the LBD in the antagonist-bound crystal structure precludes the LBD from sampling its intrinsic fluctuations as seen in its isolated forms. Thus LBD-TMD coupled movements that are deemed important for channel opening may not be enabled in the current configuration of AMPARs.

4.2 KEY "EFFECTORS" OF ALLOSTERY IN AMPAR

The Perturbation Residue Scanning technique (PRS) introduced by Atilgan et al., 2009 (Atilgan and Atilgan, 2009) provides a metric for assessing the response of individual residues to external perturbations exerted on others. The resulting PRS matrix evaluated for AMPAR is shown in **Figure 33A**. The map provides complete information on the displacement induced at residue *j*, upon perturbing residue *i,* by a unit deformation, for all $1 \leq i, j \leq N$, where N is the total number of residues (3116 in AMPAR, 779 resolved residues in each chain) (see Methods). The PRS permits us to assess the residues distinguished by their strong *effector* or *sensor* properties. Effectors elicit a high perturbation response, as evidenced by the peaks along the *right ordinate bars* in **Figure 33A**, each bar being obtained by averaging the elements in the corresponding row. Likewise, the averages over the elements of each column yield the sensor propensity profile (*lower abscissa*), where the peaks highlight regions that are most *sensitive* to perturbation.

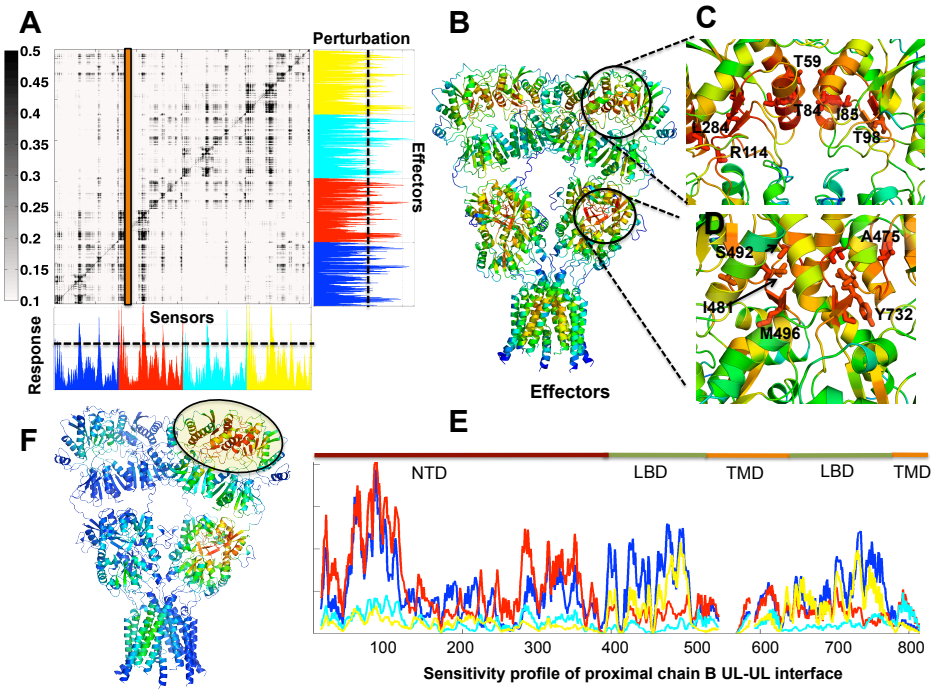


Figure 33. Illustration of sensitive and highly influential regions of AMPARs using PRS.

(A) Normalized PRS matrix where strongest perturbation-response regions are shown by in *black* (see scale on the *right*). Bar plot along the lower abscissa shows the average response to perturbation (average over rows of PRS matrix) and that along the right ordinate shows the average propensity to trigger perturbation (average over columns in the PRS matrix). Each bar is colored coded by chain identify as in **Figure 29A**. The peaks along the bar plots indicate the effectors (*right ordinate*) and sensors (*lower abscissa*). **(B)** AMPAR structure color-coded by ability to influence perturbation (perturbation shown in bar plot along *right ordinate*), where *red* regions are peaks in the bar plots. **(C-D)** A zoomed-in view of the "effectors" in the NTD **(C)** and the LBD **(D)** highlighting some the key residues that show high efficacy in inducing perturbation. **(E)** The sensitivity profile (panel 2 *orange* bar) of T84 in the proximal subunit B. The peaks in each chain represents residues whose perturbation strongly couple with $T84_B$. **(F)** AMPAR structure color-coded by sensitivity profile, with ellipse highlighting the LBD region that affects perturbation at $T84_B$ of NTD and its neighboring regions.

Figure 34A shows the AMPAR structure color-coded by the propensity of residues to serve as sensors, with *red* regions indicating high responses to perturbations, and *blue* regions, low responses. Note that exposed regions (aA and aI helices, and top loop regions) in the NTD upper lobes, and the intracellular termini of TMD helices are distinguished by their high sensitivities. On the other hand, **Figure 33B** highlights the strongest effectors (in *red*). We

observe that these residues are located in the core regions of the domains. A zoomed-in look at the "effectors" in the NTD and LBD is shown **Figure 33C** and **D,** respectively. Interestingly, in the NTD, these residues are located at (i) the NTD dimer interface (e.g., T59, T84 and I85), (ii) the NTD cleft that facilitates clamshell-like closure (R114) and (iii) the region connecting these sites. A similar pattern is seen in the LBD, where residues near the ligand-binding site and the LBD dimer interface (I481, A475, S492) emerge as strongest effectors. Their central positions confer them the intrinsic capacity to be serve as key mediators of allosteric signaling.

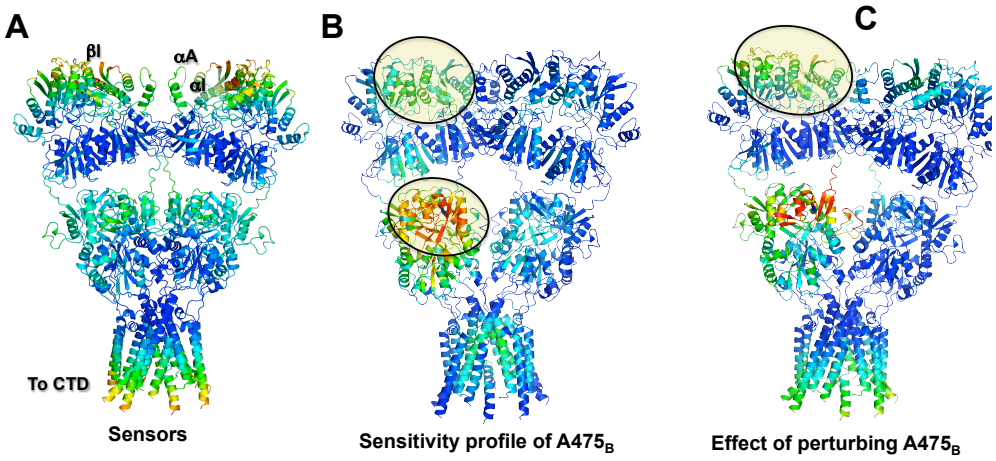


Figure 34. Coupling between NTD and LBD obtained from PRS.

(A) Shows the AMPAR structure color-coded by sensitivity to perturbation (response shown in bar plot along *lower abscissa* in **Figure 33A**). **(B)** AMPAR structure color-coded by sensitivity profile of A475 in the LBD proximal subunit B with ellipse highlighting the distant NTD region that affects perturbation at $A475_B$ of LBD and its neighboring regions. **(C)** AMPAR structure color-coded by displacement/response profile of all residues as a result of perturbation at $A475_B$ of LBD. The ellipse highlights the distant NTD region that is affected by perturbation at the LBD.

Figure 33E shows the perturbation sensitivity profile of one of the key "effectors" located at the UL-UL NTD interface (T84 proximal subunit *B,* **Figure 33A**, highlighted bar). Residues (in different chains) whose perturbation has a high influence on this key residue appear

as peaks in **Figure 33E** and as *red* regions in **Figure 33F** showing the profile mapped on the AMPAR structure. Interestingly, apart from regions near T84 UL-UL interface, there is evidence of strong coupling near the ligand binding and interfacial regions of the LBD AD dimer (circled ellipse in **Figure 33F**). Weaker associations also exist with the NTD *CD* and LBD *BC* dimer, continuing to the TMD. Such couplings between the NTD and LBD also exist for effectors at the LBD, shown in **Figure 34B** for A475 (proximal subunit B). Also, the effect of perturbation at the LBD (A475) may be experienced in the distant ULs of the NTDs (**Figure 34C**). These observations lend support to the existence of inter-domain communication in AMPARs.

In order to understand the mechanism of signal transmission between these regions (**Figure 33 E-F**) and the TMD, we proceed next to the identification of potentially functional residues from structural dynamics and sequence evolution analyses, which may enable information propagation.

4.3 NETWORK REPRESENTATION OF PROTEINS HELPS IDENTIFY MEDIATORS OF INFORMATION FLOW

Allosteric communication is essential to function, especially in multimeric proteins. Allosteric signal transduction is usually accepted to be mediated by key residues distinguished by their high "allosteric potential" (Bahar et al., 2007; Ming and Wall, 2005). Conserved residues (Suel et al., 2003), allosteric binding sites (Liu et al., 2007), regions that support large-scale domain motions (hinge sites) (Dutta and Bahar, 2010; Liu et al., 2010; Yang and Bahar, 2005) may all hold high propensity for proficient signal transduction. Network models of proteins and graph theoretical approaches enable us to identify such sites that play a critical role in information transfer. We

have utilized a series of such network-based approaches, summarized below, to deduce consensus sites potentially implicated in allostery.

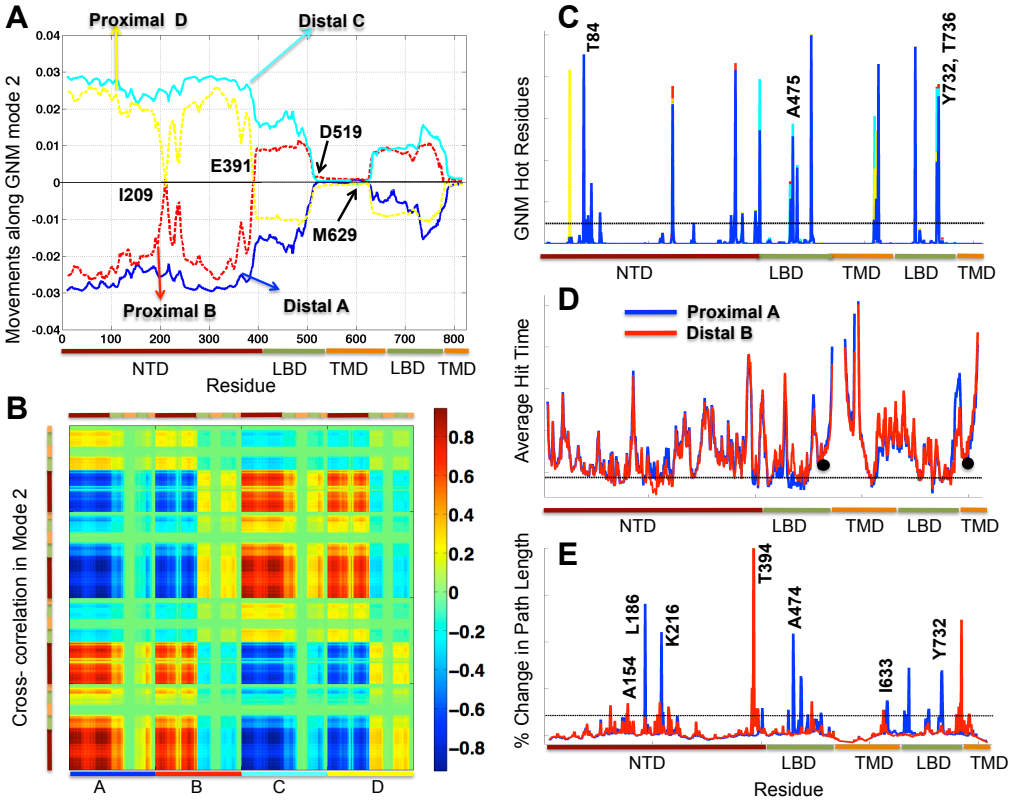


Figure 35. Residues with high allosteric potential identified using network models and graph theoretic approaches.

(A) Shows the distribution of motions in the second slowest mode of the GNM of each chain (color-coded as in **Figure 29A**), with the important hinge residues shown explicitly. **(B)** The cross-correlation map in slow mode 2 of GNM showing regions that are strongly coupled (*dark red* and *deep blue*) (see scale on *right*) within chains and between chains. **(C)** The average of the fastest 40 modes from GNM is plotted for all chains. The dotted line indicates the cutoff above which the residues were identified as "hot spots". The hot spots are identical for almost all the residues (exact overlap of all chains). **(D)** Shows the average hit-times plot for one representative distal (*blue*) and proximal chain (*red*) each. The dotted line indicates the cutoff below which the residues were identified as "fast signal propagators, with two additional residues shown in black markers. **(E)** Shows change in characteristic path length as the residues in one distal chain (*blue*) and one proximal chain (*red*) are knocked out of the structure. The dotted line indicates the cutoff above which the residues are considered to be "central" to the topology, since they change the path length appreciatively.

4.3.1 Hinge residues and kinetic hot spots deduced from the GNM

The GNM (Bahar et al., 1997; Haliloglu et al., 1997), a widely used method for characterizing the equilibrium fluctuations and cross-correlations of proteins, provide us with information on two types of mechanically significant residues: *hinge residues* at the interface between protein substructures (domains/subunits) that undergo concerted movements, and *kinetic hot spots*, or 'centers of energy localization' that confer strong resistance to structural deformation (Bahar et al., 1998; Demirel et al., 1998). The former are deduced from the most constrained anchor-like regions of the global (lowest frequency) mode profiles (**Figure 35A**), and the latter, from the peaks of the highest frequency modes (**Figure 35C**).

As may be seen in **Figure 35A**, a highly collective global mode (GNM mode 2) partitions the two proximal subunits into two regions that undergo opposite direction movements, separated by a hinge site at the interface between the NTD and the LBD, centered around E391 (see **Table 6**). Additionally, the entire TMD appears to be highly immobile in this mode. The LBD/TMD linker residues, D519 and M629, serve as hinges at the flexure/bending region between these two domains. We also note that I209, at the interface between the two NTDs appears to be highly constrained. Notably, the global mode shape captures the swapping of domains where each proximal chain links with a different distal chain as it moves from NTD to the LBD. The cross-correlation map associated with this slow mode (**Figure 35B**) highlights the correlated (*red*) and anticorrelated (*blue*) regions across the tetrameric structure. Note that the TMDs do not show significant motion/coupling with other domains (*green*).

In **Figure 35C,** on the other hand, the identities of the peaks (hot spots) are maintained across the subunits (the curves for the four subunits overlap). This is due to the fact that the high frequency modes refer to localized fluctuations; they basically involve the highest packing

density regions within the individual domains, Note that several peaks coincide with the effectors identified above by PRS.

4.3.2 MSM analysis reveals fast information propagators

A discrete time, discrete state Markov model has been shown to describe the stochastics of signal propagation in proteins (Chennubhotla and Bahar, 2007) (Chapter 2, section 2.2). An important metric that MSM yields for individual residues is their *Hit Time, H(i,j),* which quantifies the total number of steps it takes to transmit a signal from residue *j* to residue *i* averaged over all possible paths. *H(i,j)* is in turn used for calculating an average hit time <H(*i*)> for each residue, averaged over all others residues, $i \neq j$. **Figure 35D** shows the <H(*i*)> profile for representative distal (*blue*) and a proximal (*red*) subunits of AMPAR. The dashed line indicates the upper cutoff chosen for identifying residues distinguished by their short hit times, i.e. fast signal-propagation properties, and black dots indicate additional residues that have minimal hit times at the LBD/TMD linker. See **Table 6** for the identity of those residues.

4.3.3 Central residues

Network models for proteins exhibit small-world network properties (Atilgan et al., 2004), and as such, they have *central residues* vital for internode communication (Atilgan et al., 2004; Watts and Strogatz, 1998). In this residue network model, two nodes are connected if at least one atom of one is within 4.5 Å distance from any atom of B, and the smallest distance between connected atom pairs represents edge weight. A robust measure of graph topology is the characteristic path length (L) defined as the average shortest path length between all pairs of

nodes. A residue *i* is a "central residue" if its removal from the network causes a significant change in L, or a large $\Delta L_i = L_{new(i)} - L$ (del Sol et al., 2006). This criterion has been used to identify functional residues like catalytic sites, mediators of protein folding, and substrate-binding hotspots (del Sol et al., 2006). The peaks in **Figure 35E** show the central residues in AMPAR distal (*blue*) and proximal (*red*) chain.

4.3.4 Residues with high "allosteric potential"

Before we proceed to the integration of results, we also evaluated the evolutionary conservation of residues, given that sequence variation is tightly coupled to structural dynamics (Liu and Bahar, 2012; Marks et al., 2011; Worth et al., 2009). We determined conserved residues based on scores from the Consurf server (Ashkenazy et al., 2010) for multiple sequence alignments (MSAs) of the NTD and LBD obtained from the Pfam database (Punta et al., 2012) (see *Appendix C* for details). A compilation of the results from the above analyses (effectors, fast signal commuters, hinges, hot spots, central residues, conserved residues) is presented in **Table 6**. Overall, we identified 280 residues across the tetramer (a total of N = 3116 residues, 779 residues per subunit) that are suggested to play a role in mediating allosteric communication.

Table 6. Key residues obtained with different network-based methods and conserved sites

		Fast Signal Propagators	Effectors PRS	GNM Hinges [(a)]	Central Residues	GNM Hot spots	Conservation
Both / Proximal / *Distal* subunit	NTD UL		**T59, T84, I85, T98/** G75, S95, S95, R114, Q112, L284			**S81, T84/**S55/ *S95*	**G75, K79, S95, T98, Q112, R114, D287, A288, V289, T293, A295**
	NTD LL	**W135, Y137, Y218/** V205, Q207, T210, K213, H214, F245/*R190*		**E391/**T383, *E384*	L122, A154, K213, H214 T383, T386, T394/*L186, K216*	**F245, I360 M361,W374**	**L117, A120, A223, W130, E241, I481, R485**
	LBD S1	**F491, Y405, Y732, Y424/** *V396, T398, A429, A475, L479, I481, R485, M463, A734*	**A475, F491, S492, P494, L479, Y732, I481, I734/** R480	Q392, T394	*I459, W460, D473*	**T398, L467, A475, Y732, T736, V406, M463**	**G423, D473, D490, F491, S492, N747**
	LBD S2	**W766**		Q509, F515, M625, T625/ *I508, P632*	*N709, M629, P632, I633*	**S501, V693**	**Y700, L704, W766**
	TMD	**A618, A622**		**K781/** A622, S788		**A618, A621**	

Residues identified as fast signal propagators, effectors from PRS, GNM hinges, central to topology, GNM hot spots, and conserved residues in distinct regions of the AMPAR structure. We divide the results into a) those identified for both chains (in *bold*), b) those only for proximal (in *italics*) and (c) those only for distal chains. The residues are not unique for each method and overlap across different methods. Overall we identify 280 residues across 4 different chains (68 residues for each of the two distal chains and 72 residues for each of the two proximal chains).

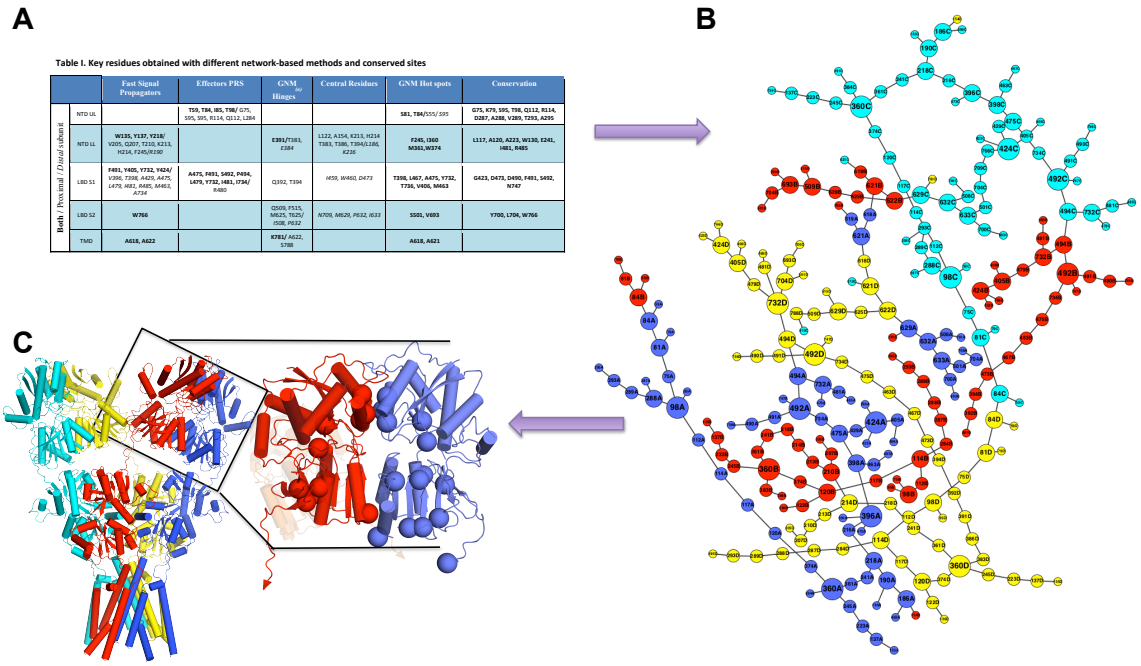


Figure 36. Illustrating the protocol adopted to ascertain an allosteric communication pathway.

(A) Shows a table of residues identified as "key allosteric mediators" using network properties illustrated in Figure 3 (See Table 6 for the complete list). A sub graph containing these residues (total 280 residues from all chains) as nodes is constructed and two nodes have an edge if their C^a distance <15 Å (this cutoff was chosen to form a connected component of the sub graph), with C^a distance as edge-weights. A minimum spanning tree of the sub graph is then calculated using Prims algorithm and is shown in **(B)**. The nodes are colored according to their chain identity (as in Figure 29A), and the sizes are in line with the node degree in the tree. This gives the most effective signal propagation network. **(C)** Representative nodes, especially ones with high degrees are mapped onto the structure.

Next, we integrate the data in **Table 6** to infer the allosteric pathway(s) that potentially modulate AMPAR function. **Figure 36** is a pictorial depiction of the steps used to this aim. We

use the data in **Table 6** to construct a fully connected graph, the nodes of which are the tabulated residues, and the edges connect those pairs with C^a- C^a distances are <15 Å. Edges are assigned weights based on their C^a- C^a distances. Using the minimum weight-spanning tree (Prim, 1957), we obtain a minimally connected sub graph (**Figure 36B**), with smallest edge-weights or most efficient information transfer properties. The nodes/residues with highest number of edges (highest degree nodes) form hubs in the sub graph, depicted by bigger node size. **Figure 36C** shows the next step in the method where we map the residues onto the structure.

4.4 NTD MEDIATED ALLOSTERIC PATHWAY IN AMPAR

Figure 37 displays the spatial distribution of key residues identified by this integrative methodology. The arrows indicate the paths of information transfer and the interfacial residues that enable this transfer. Interfacial residues include those pairs that make intradomain (e.g. UL-UL, LL-LL, UL-LL, S1-S1, S1-S2) or interdomain (NTD LL to LBD S1 and LBD S2 to TMD gate) contacts, as summarized in **Table 7.** A few highlights are the following:

(i) T84 is particularly imported in the transmission of signals across the monomer-monomer interface of NTD ULs (**Figure 37A)**. Mutation of this residue have been observed to destabilize the homodimer interface (Rossmann et al., 2011), thereby lending support to the significance of the proposed network

Table 7. Residues Implicated in Interlobe, Interdomain or Intersubunit Communication

Table 2. Residues Implicated in Interlobe, Interdomain or Intersubunit Communication				
NTD		Residues[(a)]	Chains	Color index
UL - UL Dimeric Interface		{**T59, T84, S81**, K79}$_A$ → **T84**$_B$	*A-B or C-D*	*Orange*
UL – LL Communication		G75, **T98**, Q112, **R114**, L117, **A120**, W374, **I360**	*within A, B, C or D*	*Pink*
UL-LL monomer interface to LL-LL dimer interface	Path 1	{**(I360), F245, A223, Y137, W135**}$_A$ → **Y137**$_B$	*A-B and C-D*	*Blue*
	Path 2	{**(I360), M361**, **E241**, **Y218, R190, L186**}$_A$ →**A154**$_B$	*A-B and C-D*	*Yellow*
	Path 3[(c)]	{**I360 → M361**, **E241**, **Y218, H214**}$_B$ → {**T210, K213**}$_D$	*B-D*	*Yellow*
NTD LL – LBD interface		**I360 → M361**, **E241**, **Y218,** K216, **V396**	*within A or C*	*Yellow/ Green*
		I360 → T383, **T386**, **E391**, Q392, (**T394)**	*within B or D*	*Green*
LBD		**Residues**	**Chains**	**Color**
NTD interface – LBD S1 core		{(**K216**), **V396, T398, A475,** I734, (**S492**)}$_A$	*within A or C*	*Green*
		{(**T394**), **D473,** L467**, M463, A475,** I734, (**S492**)}$_B$	*within B or D*	*Green*
S1-S1 Dimer Interface		{(**S492**), **P494**}$_A$ → {**P494,** F491, D490}$_D$	*A-D or C-B*	*Orange*
S1-S1 interface to S1 Ligand-Binding Region		(**S494**), **Y732,** L479, **I481,** R485[(b)]	*within A, B, C or D*	*Orange/Blue*
S1 → S2	Path 1	{(**A475**), **A429, Y424, W766**}$_A$	*within A or C*	*Green/Pink*
	Path 2	{(**Y732**), L479**, Y405, Y424, W766**}$_B$	*within B or D*	*Blue/Pink*
	Path 3	{**Y732, L704**}$_B$	*within B or D*	*Blue*
S2 → LBD/ TMD Linker	Path 1	{(**N709**), **L704, S501, I633, P632, M629**}$_A$	*within A or C*	*Blue/Green*
	Path 2	{(**L704**), **V693, Y700, K509, M629, T625**}$_B$	*within B or D*	*Blue/Green*
LBD-TMD Linker/ TMD Gate		{(**M629**)}$_A$→{**622, 621, 618**}$_D$,{**621, 618**}$_A$,{**621, 622**}$_B$ ←{(**M625**)}$_B$	*within A, B, C or D*	*Yellow*

(a) Subscripts indicate example chain identifier; residues in bold are displayed in Figures 6 and S3.
(b) Coordinates ligand

(ii) I360 on LL, in proximity to UL-LL interface, forms a hub, that channels information in various directions to effect distal-proximal, proximal-proximal and interdomain (NTD-LBD) communication (see also **Figure 37B**).

(iii) The close proximity of NTD K216, R190 and L186 with respect to the LBD S1 in the distal chains, which is due to the compact nature of the NTD/LBD linker, constrats that at the linker (or interdomain hinge regions – T386-T394) of the proximal subunits. In the latter case, the more distant and flexible positioning of the NTD LL and LBD S1 results in a loss Markovian information transfer (Chennubhotla and Bahar, 2007). Thus the bulk of the signal transmission across the intact AMPAR apparently occurs through the distal chains. This property is also supported by the view that conformational changes in the distal

subunits contribute more towards channel opening than those in the proximal ones (Das et al., 2010; Sobolevsky et al., 2009).

(iv) A475 is instrumental is transferring information from the NTD-LBD interface to the LBD S1-S1 interface, specifically to S492, P494 (*orange*) that form strong hubs in the network, and reach out to glutamate-coordinating residues P478, T480 and R485 (Pohlsgaard et al., 2011) (**Figure 37C**). This communication may be significant, since rearrangements in LBD dimer interface have been implicated in de-sensitization (Armstrong et al., 2006). Furthermore, in NMDARs, rearrangements in the LBD dimer interface have been proposed to modulate the allosteric signals originating from the NTD (Gielen et al., 2008; Mony et al., 2009).

(v) Two pathways appear to establish LBD-TMD information transfer: As shown clearly in **Figure 37C-D**, these two pathways originate from LBD S2, both proximal and distal chains to merge at T625 near the TMD gate, where a tight interaction with gating residues (A622, A621, A618, *yellow*) is distinguished. These strongly interconnected residue network controls the opening of the TMD gate.

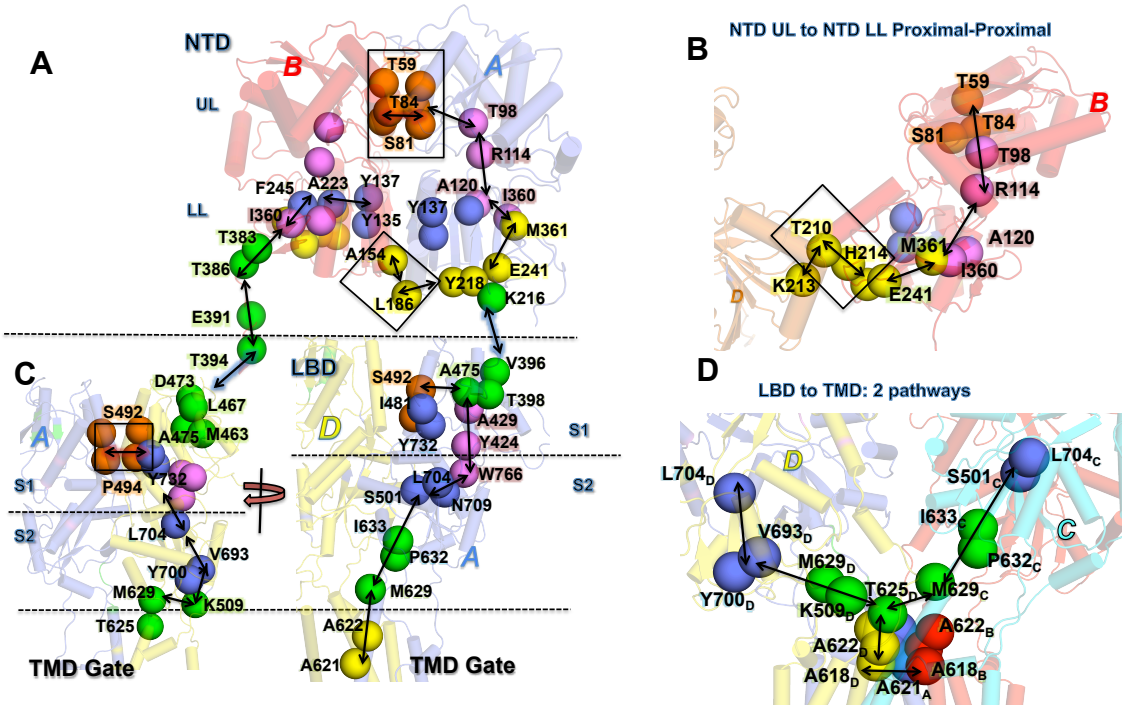


Figure 37. Mapping the NTD mediated allosteric pathway in AMPAR and proposing a hypothetical mechanism of control.

(A) Shows representative residues from Table 1, mapped onto an NTD distal-proximal pair (*AB*) as spheres, with arrows showing the dissipation of information from NTD UL-UL interface (box), via NTD cleft to LL interface, and NTD-LBD interface. **(B)** Shows inter dimer information propagation between the proximal chains of the NTD indicating a path from the NTD UL interface, via cleft to the LL, and then channeled by $\{E241, H214, T210\}_B$ to $K213_D$ of the other proximal chain. **(C)** Similar as in A, showing information transfer from the NTD to the LBD S1 (*AD* dimer), along the proximal (right) and distal chains (left) to the LBD ligand-binding region, dimer interface, S2 domain and further downstream towards the TMD **(D)** Highlights the two pathways of information transfer (in proximal and distal chains) from the LBD reaching an interconnected network of residues in the TMD. See also **Table 7**.

Figure 38 is a cartoon representation of the proposed allosteric response to ligand binding at the NTD that follows from the connections established in our network analysis described above. Ligand binding at the NTD cleft (R114) could affect changes at the UL dimer interface as well as the LL (A120, I360), and eventually propagate the signal to the LBD. In the

LBD, this would affect glutamate binding and the conformation of the subunits at the dimer-interface, thereby propagating the signal to modulate TMD gating.

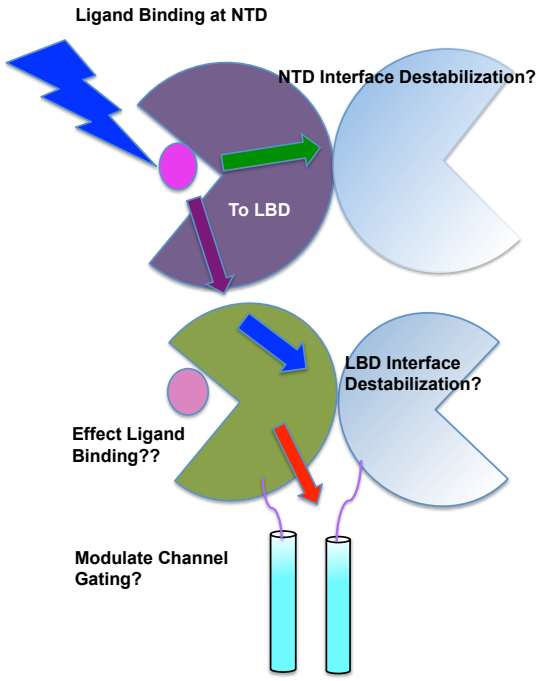


Figure 38. A proposed mechanism of allosteric signaling in AMPARs.

Cartoon representation of a probable mechanism of allosteric control by AMPAR NTDs based on observations from network analysis.

4.5 DISCUSSION

Resolution of isolated LBD and NTD structures has enriched our understanding of the iGluR's domain architecture, dynamics and responses to ligand binding (mainly LBD). However, the pivotal point defining our understanding of the functional (tetrameric) organization of iGluR domains has been the determination of the crystal structure for the full-length GluA2 AMPAR (Sobolevsky et al, 2009). This structure has enabled us to make inferences on the architecture of non-AMPARs, such as kainate- and NMDA-receptors. Many recent studies elucidated that these

closely resemble AMPAR's (Das et al., 2010; Lee and Gouaux, 2011; Rambhadran et al., 2010; Salussolia et al., 2011; Sobolevsky et al., 2009). Here we make a further step, elucidating the structure-based dynamics of the tetrameric receptor, which has not been studied to date.

In this study, we provide a first glimpse into the global mechanics of the AMPAR tetramer, which discloses a significant cooperativity between the movements of the different domains. Mainly, the NTD dimers' ULs undergo large fluctuations resulting in their 'distant' and 'close' positioning (**Figure 29B** and **29D**), while the LBD undergoes a coupled global twist, accompanied by the rotation of the TMD around its central axis of 4-fold symmetry. Given the universality of the tetramer assembly in iGluRs, and based on the fact that global motions are encoded by the overall structure, one might anticipate that such a global motion may be favored in other iGluRs as well.

We also show that packing of the NTD in the full-length receptor does not inhibit the intrinsic motions of the individual dimers. The modes previously characterized in isolated NTD dimers suggested that an allosteric potential akin to NMDARs (Dutta et al., 2012), and the present study shows that these are likely to be retained in the intact receptor. The same property is not true, however, for the LBD in the currently resolved conformation. In principle, the isolated LBD monomers undergo different degrees of cleft closure due to their bilobal architecture (**Figure 30-31**). In the present crystal structure, the tight packing of the LBD monomers against each other, and the close association between the LBD and the TMD obstructs the accessible movements, leading to alterations in the dimer dynamics (compared to isolated dimer). The NTD dimer, on the other hand, can undergo its clamshell-like as well as monomer-monomer counter-rotations (shearing movements), suggesting that it maintains its sensitivity (and response) to perturbations. **Figure 32** further shows that perturbations at the UL-UL

interface, that may originate from ligand-binding, have an impact on the LBD S1, this effect being particularly pronounced when perturbing proximal subunits. Not surprisingly, in both the NTD and the LBD, the PRS analysis indicates that regions in the dimer interface and cleft region appear as being key "effectors" of allostery. The cleft motion has been suggested to influence ligand binding of allosteric modulators in NMDA receptors (Dutta et al., 2012; Karakas et al., 2011; Sirrieh et al., 2013), and may play a similar role in the AMPARs. Rearrangements of the LBD ligand-binding domain and the dimer interface are already established features that modulate iGluR function (Armstrong et al., 2006; Furukawa et al., 2005; Gielen et al., 2008). Next, we identified residues that may allow information propagation between the "effector" regions of the NTD and LBD.

Lastly, based on graph theoretic analysis, we proposed that the most efficient and probable NTD mediated pathway of allosteric communication would stem from NTD UL interface (T84), through the cleft (R114), to the NTD LLs'. Residue I360 forms a hub that channels information through more than one pathway. One notable feature of the pathway leading to LL interface of the dimer is Y137 that projects into the interface. In GluK2, the cross-linking of inter-monomer L151C affects activation by glutamate (Das et al., 2010). Y137 is close to the AMPAR counterpart (L143) and may be a suitable candidate for testing the effect of the AMPAR NTD on channel-gating properties. Cross-linking at the tetramer interface in GluK2 (G215) also affects activation albeit to a lesser extent than the links at the dimer interface (Das et al., 2010). Regions near the GluA2 counterpart (G212, namely H214, T210) are involved in inter-dimer communication across the proximal subunits, also suggesting their role in signaling.

The signal is transferred from the NTD to the LBD along the distal NTD/LBD interface, and along the linker for proximal chains. Whereas, there are concerns regarding the existence of

the NTD/LBD interface due to missing six residues of the linker, recently there have been suggestions of the downward pulling of the NTD towards LBD and shortening of the length of the receptor in NMDARs (Borschel et al., 2011; Sirrieh et al., 2013). Also noticeable from ENM analysis of AMPARs is the flexibility of the linkers. From these observations, the distal NTD/LBD interface may be accessible even in the presence of the missing 6 residues of the linker.

This signal is then transmitted through some residues of the S1 domain of the LBD (D473, M463, A475) to the region near the LBD ligand binding and the LBD dimer interface (S492, P494, Y732, I481). Furthermore, signal is then transmitted via Y424, A429 and F766 to N709, L704 in the S2 domain. Some of the residues near glutamate-binding region of AMPARs (L704, Y732) are NMDA coordinating residues in the NR2A counterpart of the LBD (Furukawa et al., 2005). The role of H-bond between W766-D727 in cleft closure has also been recently elucidated (Ahmed et al., 2013). This suggests that their dynamics may play a role in ligand binding even though they do not coordinate glutamate in the crystal structure of GluA2. From the LBD S2, the signal is propagated via S501, I633 and M629 (also V693 and K509) to the TMD gate (A622, A621, A618) (Sobolevsky et al. 2009).

Thus we hypothesize a mechanism of interdomain communication in AMPARs and propose a network of residues that mediate this pathway. Some of them have already been established experimentally as being important for function in iGluRs, while others may be tested for putative role in function. In fact, the signaling mechanism that we characterize closely follows the signal transduction cascade involving inhibition by Zn^{2+} that is established for GluN1/GluN2A NMDARs involve (Gielen et al., 2008; Mony et al., 2009). Based on our

observations, a similar mechanism cannot be ruled out for AMPARs. Thus, we open newer avenues of probing the NTD mediated allosteric mechanism in AMPARs.

The results from this chapter are being organized into a manuscript: Dutta A and Bahar I. *A series of residues acting as sensors and effectors regulate allosteric communication in the intact AMPA receptor.*

The studies describing the putative allosteric potential of AMPAR NTDs leaves an open question whether the NTDs are in fact "druggable". To this end, we use an MD simulation based exploration of the ligand binding landscape of iGluR NTDs that enables the incorporation of conformational flexibility of the target and probe molecules and provide more realistic detection of binding sites and the maximal binding free energy (Bakan et al., 2011; Seco et al., 2009).

5.0 DRUGGABILITY OF IGLUR NTD AND LBD

Failure of drugs passing toxicity/efficacy measures especially in later stages of development has resulted in a very steep increase in costs of development of new drugs. Thus, it is important to choose targets that are in fact "*druggable*" (Hopkins and Groom, 2002; Keller et al., 2006), similar to choosing ligand that have "*drug-like*" (Lipinski et al., 2001) properties in early stages to avoid attrition effects arising at later phases of drug development. This has led to an increased focus on structure-based drug design for early consideration of molecular interactions between target and ligand to foster design/selection of more potent and selective molecules. Screening of fragment-sized compounds or organic probe molecules that seem to occur frequently in drug molecules via experimental methods like NMR and X-ray crystallography has proved highly useful in making assessments of binding hot spots that contribute maximally to binding free energy (Allen et al., 1996; Hajduk et al., 2005). But such methods are expensive, thus giving rise to a need for more cost-effective computational mapping methods. Surface descriptor based methods (Cheng et al., 2007; Nayal and Honig, 2006), docking of fragment-sized molecules (Brenke et al., 2009; Huang and Jacobson, 2010) are a few techniques that have been developed with reasonable predictive capacity for druggable sites. Their successes however, have been limited to more "rigid" binding pockets. Thus there is a need for incorporating target flexibility: large-scale dynamics (Bakan and Bahar, 2009; Floquet et al., 2006; May and Zacharias, 2008) as well as side-chain and backbone reorientations due to thermal fluctuations (Brown and Hajduk,

2006; Ivetac and McCammon, 2010) for more accurate modeling of flexible binding sites and identification of allosteric sites.

The use of MD simulations of target proteins in water and organic probe molecules has been developed recently to bypass some of the above limitations (Guvench and MacKerell, 2009; Seco et al., 2009). More recently, Bakan et al., 2012 (Bakan et al., 2012) have used an optimized probe set that captures fragments over-represented in existing FDA approved drugs for assessing simulation based assessment of probe binding. They show that this rigorous approach can successfully predict binding sites of a wide variety of targets including allosteric sites. Thus, MD simulation based "druggability" assessment are more time-consuming (than docking techniques) but offer a rigorous accounting of various molecular driving forces, including entropic and desolvation factors and allow alternate conformation sampling with a potential for opening new druggable sites (Bakan et al., 2012). This gives us an opportunity to explore the binding landscape of iGluR NTDs, especially for AMPARs, for whom no known ligand binding to the NTD has been established till date. First we provide a brief description of the methods.

5.1 BRIEF OUTLINE OF METHODS

The details of this pioneering method can be found in Bakan et al., 2012 (Bakan et al., 2012). Simulations were performed using NAMD (Phillips et al., 2005) using CHARMM force field (MacKerell et al., 2002) and CHARMM general force fields (Vanommeslaeghe et al., 2010). System setup, equilibration, and productive simulation protocols were performed as described by Bakan et al., 2012 for 40-60ns of productive run (See Appendix C for details regarding the system used). Despite the short time scales, it has been shown that due to fast diffusion of small

probes, this method can successfully explore entire protein surface, locate and bind protein pockets, and induce structural changes akin to that seen in the bound form (Bakan et al., 2012). The estimation of free energy is based on the premise that simulations of 40-60ns will result in a Boltzmann distribution of the probe molecules, which enables us to convert probability distribution into binding free energy. Next a grid-based approach is adopted to find the density of probes in each grid and compare it with the expected density. This results in the calculation of binding-free energy at each grid *i* ΔG_i probe-binding = $-RT \ln(n_i/n_o)$.

Here n_i / n_o is the ratio of the observed density of probes n_i, to the expected density n_0, R is the gas constant and T is the absolute temperature (in kelvin). This free-energy map then yields interaction spots that are identified by removing the voxels that overlap with the lower energy voxels, and each spot then represents a volume equal to that of probe molecules. 7,8 proximal (located within 5.5 to 6.5Å of each other) interaction spots (28-32 heavy atoms) are then clustered and merged to locate distinct binding sites. For each binding site, maximal achievable binding affinity is calculated under the assumption that the free energy of binding calculated for the interaction spots is additive. Thus the estimation of maximal affinity for each distinct site is the sum of free energies of interaction spots in that region. Once all sites are evaluated, the most druggable site is the one that yields the lowest binding free energy.

Note in all figures (Chapter 5) probes are reported by their center of mass (shown as spheres) and color-coded based on colors shown in Appendix C. The seven main probes used are isopropanol (IPRO), isopropyl amine (IPAM), Benzene (BENZ), acetate (ACTT), acetamide (ACAM), imidazole (IMID) and isobutane (IBUT).

5.2 RECAPITULATION OF BINDING SITES OF IGLUR LBD

The LBD of iGluRs is a bilobate structure resembling the bacterial periplasmic-binding proteins that have been evolutionary optimized to capture ligands in the cleft of the lobes (Sukumaran et al., 2011) as detailed in Chapter 4. Binding of glutamate in the LBD cleft leads to the clamshell-like closure of the bilobate structure, which initiates ion-channel opening (Armstrong and Gouaux, 2000). In this study we benchmark the method by assessing how well it can capture ligand-binding features of the LBD clamshell cleft. The iGluRs LBD is a good system for this analysis because of the existence of a large number of crystal structure bound with agonists, partial agonists, antagonist (Pohlsgaard et al., 2011), particularly for GluR2 AMPA LBD.

5.2.1 Acetate binding to LBD monomer and dimers ligand binding pocket facilitates domain closure

Channel activation in iGluRs is presumably facilitated with the domain closure (S1- S2) of LBD on ligand binding and subsequent increase in the linker (P632-I633) distance between the monomers forming the dimer. This causes strain on the LBD-TMD linker region culminating into the opening of the TMD gate. The degree of domain closure is associated with the efficacy of the ligand, and affects the extent of receptor activation. We first performed simulations for the *apo* open form of the LBD monomer (two simulations with different concentrations of acetate). During both simulations, an acetate molecule interacts first with R485 that drives its entry into the ligand-binding pocket, facilitating the closure of the S1-S2 domain and forming interactions with S654 (Both R485 and S654 are known to co-ordinate ligand (Armstrong and Gouaux, 2000) (**Figure 39A**). The acetate remains bound in the pocket through most of the simulation run.

Binding of acetate is capable of triggering a domain closure (**Figure 39B** shows the degree of domain closure), but the closure is transient as the S2 domain fluctuates, even while the acetate is bound.

A similar behavior is seen in the simulations of the *apo* LBD dimer. Binding of acetate in both monomers (shown explicitly for encircled snapshot in **Figure 39** panels **D** and **E**), triggers domain closure measured by an increase in the P632A-P632B distance. Distance between linkers fluctuates between ~25 (starting) and ~36**Å**, which is similar to that seen in crystal structures with bound agonist (Pohlsgaard et al., 2011). A close-up of the binding mode of acetate in each of the dimer chains is highlighted in **Figure 39E**. In one of the subunits the acetate is bound in a similar conformation as seen in monomer simulation, interacting mainly with R485 and S654, and in a similar position as that of the carboxyl groups of glutamate and other agonists. In the other chain, however, we see acetate in a different binding position in the pocket, and its interactions are mainly with T480 (another ligand coordinating residue). Also, an imidazole probe molecule mimics the 5-membered ring-binding site of BN1 (Hogner et al., 2002) (an agonist) and forms contacts with P478.

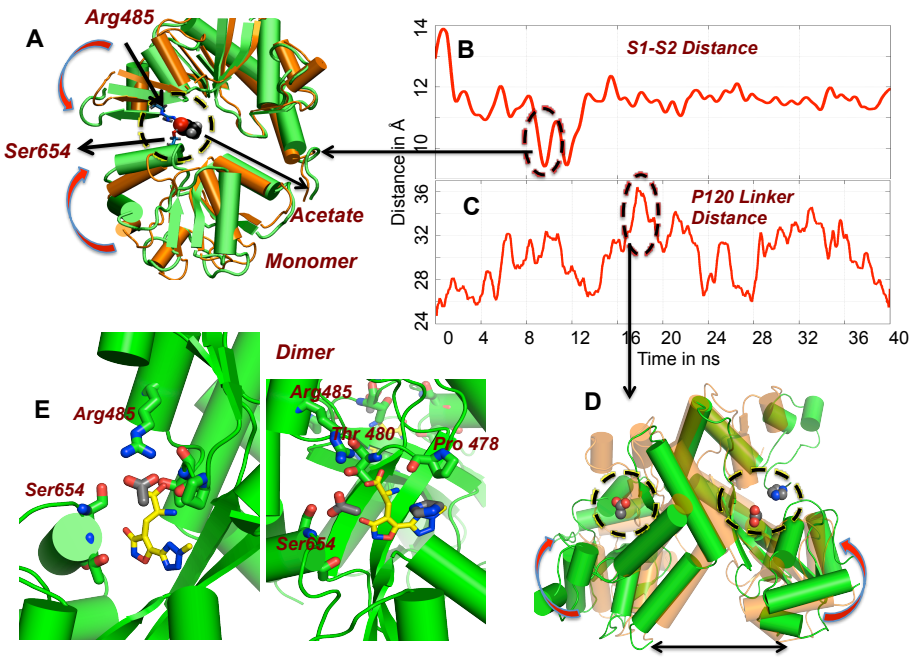


Figure 39. Acetate entry in the ligand-binding pocket of LBD monomer and dimer simulations facilitate LBD domain closure.

(A) Snapshot from simulation for the *apo* open form of the LBD monomer (1FTO-*A,* residues *5-261,* corresponding to residues *394-506, 632-773* of the tetrameric receptor 3KG2). The arrows highlight the relative closure of the snapshot encircled in **B** (*green*), from the starting conformation (*orange*) on acetate binding (shown in sphere representation) **(B):** Shows the extent of domain closure in the 40ns simulation measured by the distance between the center of mass (COM) of residues 401-403 (S1, 12-14 in 1FTO) and 654-655 (S2, 174-175 in 1FTO) (also used in Lau and Roux, 2011). **(C)** Distance between linkers, P632A-P632B in dimer simulation to measure degree of domain closure. A snapshot with high inter-linker distance (encircled in **C**) is highlighted in **D** with encircled regions showing probes bound in the ligand-binding pocket. **(E):** Shows a close-up of the probe molecules in the two chains A (*left*) and B (*right*) in the snapshot in **D** with interacting residues shown in stick representation. The bound conformation of an agonist (1M5B, Hogner et al., 2002, 2-Me-tet AMPA or BN1 with 30nM affinity) is shown in yellow stick representation. In chain *A* we see one acetate molecule (in stick), whereas in chain *B*, we see an acetate as well as an imidazole within the binding pocket.

5.2.2 LBD dimer interface allosteric modulator sites captured by probes

Receptor desensitization occurs in AMPARs by rearrangement of the dimer interface, leading to a ligand-bound, but closed-channel conformation. Whereas the exact mechanism is unknown,

crystal structures of mutant LBDs have shed some light on the structural changed associated with receptor desensitization. Positive allosteric modulation occurs by molecules like cyclothiazide (Sun et al., 2002) that stabilize the dimer interface and decrease desensitization, whereas those that disrupt the dimer-interface promote desensitization. Thus, positive modulators that enhance AMPAR currents, by decreasing desensitization have important implications in neurodegenerative diseases that cause deficiency in excitatory neurotransmission. Here we show the binding site of cyclothiazide in **Figure 40** (stick representation) at the LBD interface and the probe molecules that bind in regions overlapping this site. There is no high-affinity binding site in this region, but few probe molecules can enter the dimer interface, and capture features of the allosteric modulator binding site.

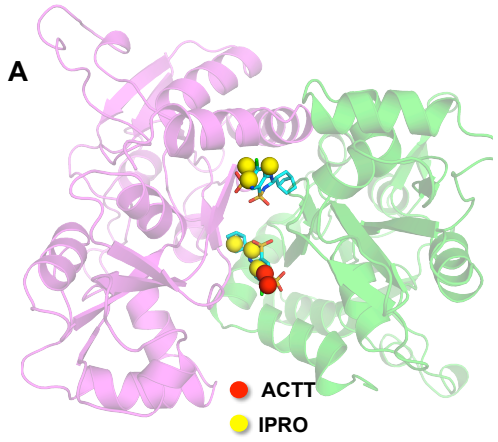


Figure 40. Probe molecules can access allosteric modulator binding sites.

(A) Shows the binding site of cyclothiazide (PDB ID: 1LBC), a known positive allosteric modulator, (stick representation) at the GluA2 LBD interface and the probe molecules (COM of probe heavy atoms represented as spheres) that bind in regions overlapping this site.

Thus, even though, we cannot estimate realistic binding affinities for the LBD ligand-binding sites (due to lack of a probe cluster), we see fragments adopt similar binding modes/sites

as that seen for agonists/modulators binding to the LBD. Next, we proceed towards capturing the binding landscape of iGluR NTDs.

5.3 IDENTIFICATION OF DIMER INTERFACES FROM NTD MONOMER SIMULATION

Among the 18 distinct subtypes that exist for iGluR (**Figure 1**), there are 9 subtypes for which crystal structures have been resolved for the NTD (**Table 8**). We performed ~40ns runs for each of the monomer subunits in the presence of probe and water molecules and the primary highlight from all 9 simulations is the identification of monomer-monomer interfacial region (**Figure 41/Table 8**). Residues that form the UL and LL dimer interface for each of the NTDs are obtained from crystal contacts with their respective homo/heterodimer counterparts (Farina et al., 2011; Jin et al., 2009; Karakas et al., 2009; Karakas et al., 2011; Kumar and Mayer, 2010; Kumar et al., 2009; Rossmann et al., 2011; Yao et al., 2011) (Dutta et al., 2012; Sukumaran et al., 2011). The only exception is the NMDA LL interface, since no distinct interface LL can be seen for NMDAR homo/heterodimers. Probe molecules that appear within 4.5Å and 5Å of these residues in the UL and the LL respectively are called "interfacial" hotspots, and shown in **Figure 41A-C** for GluA2 and reported in **Table 8** for other subtypes. For NMDA subtypes, hotspots that are close to their LL interface counterpart (as obtained from superimposition of NMDAR structures with AMPAR/Kainate) are reported. A close look at the interfacial probes at the LL interface of GluA2 is shown in **Figure 42**.

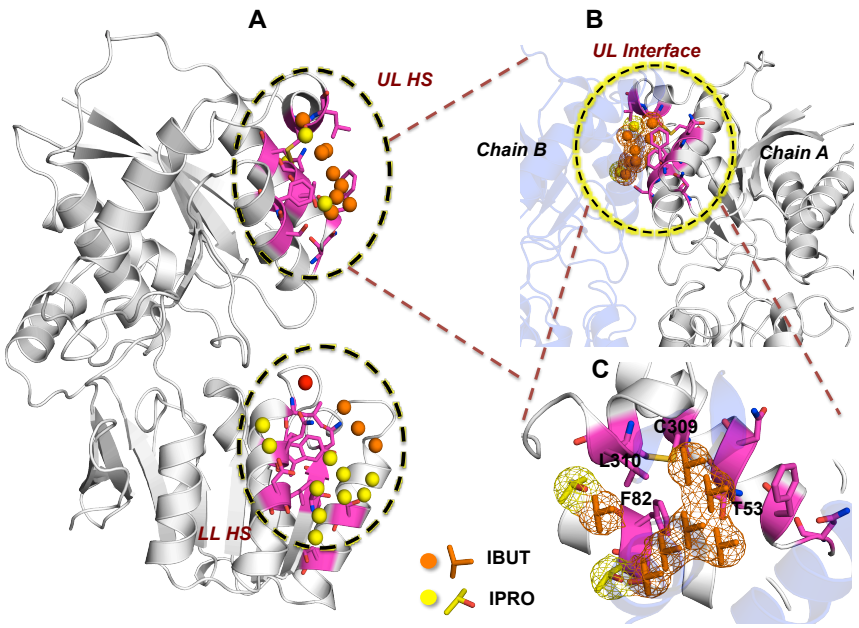


Figure 41. Protein-protein interfacial regions identified in NTD monomer simulations.

(A). GluA2 monomer with residues that form the dimer interface are shown in stick representation (*magenta*). Probe molecules that appear within 4.5 **Å** of these residues are highlighted. The spheres represent the center of mass (COM) of the heavy atoms of the probes. **(B)** We zoom in on the UL interface, where the other dimeric chain is shown (*faded blue*), highlighting that the probes indeed capture the monomer-monomer interface. **(C)** We further zoom in on the UL interface, where all the heavy atoms of the probes are shown. Note the probe configurations that are shown here, especially for probes with orientation dependent functional groups like isopropanol, are approximate (where the COM of an arbitrary probe molecule has been superimposed on the COM of the probe position obtained from simulation).

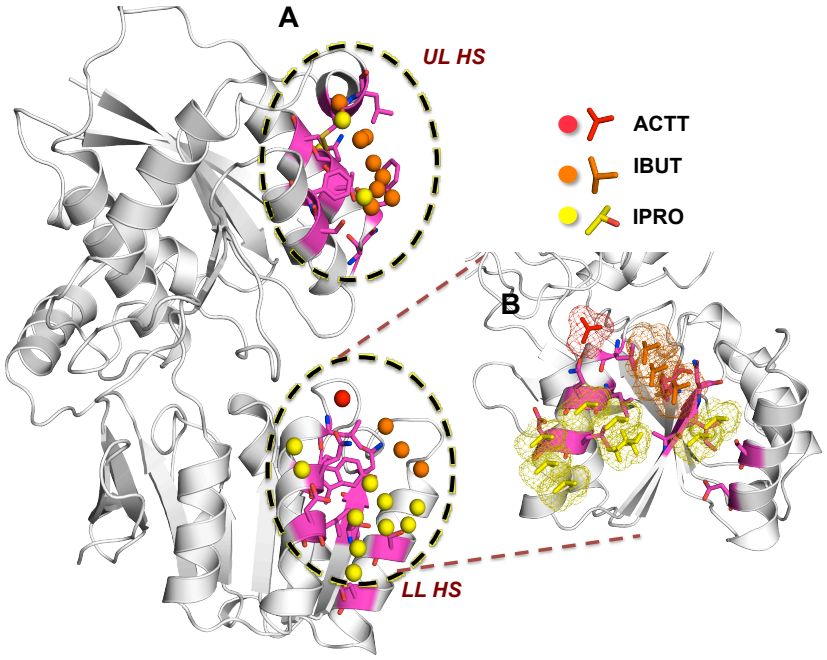


Figure 42. Probe binding at the LL interface.

(A) The LL interface residues that form the dimer interface are shown in stick representation and the COM of probes (heavy atoms) within 5Å of these residues is shown in spheres. **(B)** A closer look shows the composition and an configuration of the probes that interact with residues at the LL interface.

Table 8. UL and LL interface hot-spots in iGluR monomers.

The table reports the hotspots that appear within 4.5 and 5A respectively for UL and LL interface, and their composition. For all cases, except for NR1 LL (highlighted in red), we identify 9 or more probes that may bind at the interface.

iGluR Monomer	No of HS at **UL** dimer interface	Probe Composition **UL**	No of HS at **LL** dimer interface	Probe Composition **UL**
GluA1	23	10 IPRO, 12 IBTN, 1 IPAM	14	7 IPRO, 6IBTN, 1 IPAM
GluA2	12	4 IPRO, 8 IBTN	14	10 IPRO, 3 IBTN, 1 ACTT
GluA3	10	9 IPRO, 1 ACAM	17	14 IPRO, 2 ACTT, 1 IPAM

GluA4	9	7 IPRO, 2 IBTN	21	15 IPRO, 1 IBTN, 1 ACTT, 3 IPAM, 1 IMID
GluK2	21	12 IPRO, 8 IBTN,1 ACTT	15	10 IPRO, 1 IBTN, 4 ACTT,
GluK3	13	9 IPRO, 1 IBTN, 2 ACTT, 1 ACAM	28	18 IPRO, 9 IBTN, 1 IPAM
GluK5	9	6 IPRO, 2 IBTN, 1ACTT	14	7 IPRO, 5 IBTN, 1 ACAM, 1 IPAM
NR1	20	11 IPRO, 9 IBTN	**4**	**4 IPRO**
NR2B*	9	7 IPRO, 2 IPAM	11	8 IPRO, 2 ACTT, 1 IPAM,

**Probe abbreviations are: isopropanol (IPRO), isopropyl amine (IPAM), Benzene (BENZ), acetate (ACTT), acetamide (ACAM), imidazole (IMID) and isobutane (IBUT).*

For all cases, except for GluN1 LL, we identify 9 or more probes that bind at the interface, where primarily non-polar hydrophobic molecules like isopropanol and isobutane dominate interfacial hotspot composition. Generally, a cluster of 7-8 heavy atoms indicates a probable high-affinity ligand-binding site. However protein-protein interfaces (PPI) have larger interfaces than a ligand/protein interface and higher than 7-8 probes would contribute towards binding affinity in PPIs as seen here. Thus the above study also provides an opportunity for assessing a qualitative affinity of homodimer/heterodimer association for all NTD iGluRs. In NMDARs, the twist in the LL lobes, calls for a rather open LL interface, as seen from GluN2B-N1 heterodimer (Karakas et al., 2011) as well as GluN1 homodimer (Farina et al., 2011). Results here indicate that the LL of GluN1 subunit does not like forming an interface (fewer probes that bind at the interface highlighted in red in **Table 8**), whereas the N2B is more likely to form interfaces, even though the formation of N2B homodimer is not known. In summary, the method

consistently identifies PPI's and has future implications in identification of PPIs in proteins where interaction sites are not known.

5.4 DRUGGABILITY OF NMDA RECEPTORS

5.4.1 Highlights from NMDAR monomer simulations

Since the discovery of NMDAR-GluN2B selective compounds like ifenprodil and its derivatives, there has been a large focus on using these for differential control without competing with endogenous ligand (Mony et al., 2009, Traynelis et al., 2010). A recent crystal structure of the heterodimer of GluN1 and GluN2B, demonstrated that the phenylethanolamine binding site is located at the interface of the two NTD monomers (Karakas et al., 2011 **Figure 2B**). Another N2B modulator is Zn^{2+}, and the structure of the GluN2B subunit crystallized with Zn^{2+} showed that the binding site is at the cleft (Karakas et al., 2009, **Figure 2B**). Here, we use our simulations to see if a mixture of probes can capture the known binding sites of NMDAR NTD.

Figure 43A shows consensus probes (positively charged IPAM, from both monomer runs and dimer runs, (see *Appendix C* for table summarizing run statistics) that bind near the crystallographic Zn^{2+} binding site. There are two hot spots that interact with H127 and E284 (< 4.5Å) which coordinate the Zn^{2+} in the crystal structure (Zn^{2+} shown in violet sphere). The two probes have an affinity of ~6.4mM and their joint contribution do not account for the binding affinity of Zn^{2+} of 5.5μM as reported in Karakas et al., 2009. There is another higher affinity site comprising 2 consensus hot-spots (total affinity: 0.27mM) that bind near D102, D265 that have been reported previously to contribute towards maintaining the Zn^{2+} binding site without directly

interacting with crystallographic Zn^{2+} (Karakas et al et al., 2009). We hypothesize that there might be another Zn ion binding site in the cleft of the N2B apart from that reported in the crystal structure by Karakas et al., 2009. Overall the cleft region encompassing the above residues has a high affinity of binding +vely charged probes (4-5 total), which is captured in our simulations. Next, we evaluate the probe distribution at the polyamine binding and UL interface region of GluN2B.

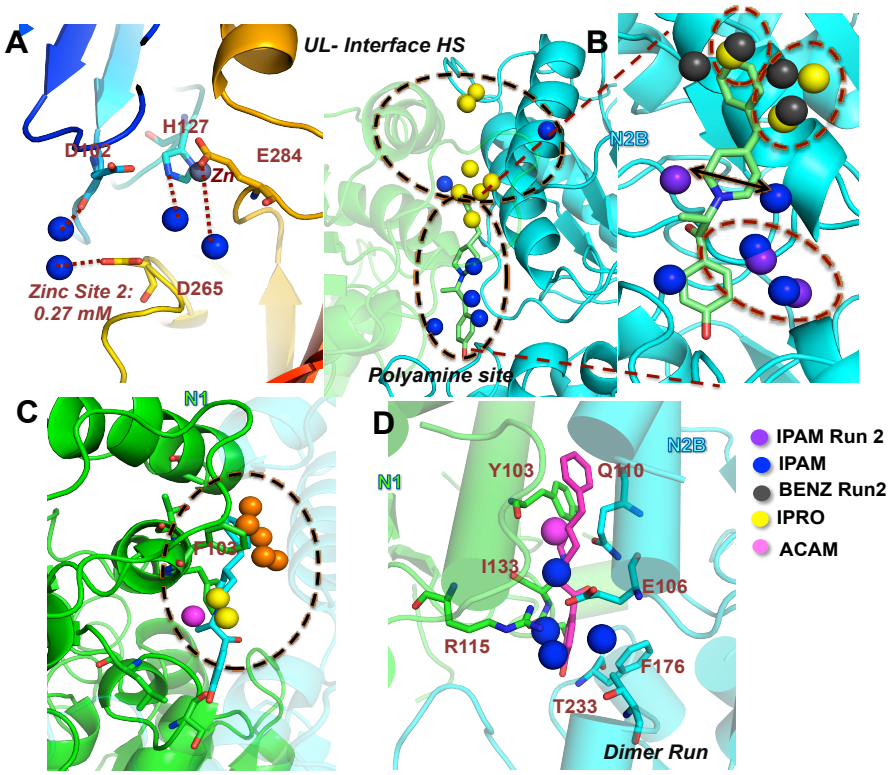


Figure 43. Capturing ligand binding sites in NMDAR receptors.

(A) Shows consensus probe binding sites (+vely charged IPAM in blue) near one of the crystallographic Zn2+ molecules (functional) (*blue-gray* sphere) in the GluN2B monomer, with the interacting residues in GluN2B shown in sticks. **(B)** In the left panel, we highlight the probes in the monomer simulation (run1) of GluN2B (*cyan)* that are close (within 4.5Å) to the UL interface and ifenprodil coordinating residues (ifenprodil shown in green stick and N1 dimer shown in faded green for reference). In the right panel, we zoom in to look at the probes overlapping ifenprodil-binding site, from both runs. Run1 identifies 7 probes (3 IPRO, *yellow* and 4 IPAM, *blue*) and Run 2 captures 7 probes (4 *BENZ*, grey and 3 IPAM, purple, some overlapping with probes from Run 1), including a common IPAM binding site near the cationic N site piperidine group indicated by arrow. **(C)** Probes overlapping the

corresponding ifenprodil binding site in the simulation of the GluN1 monomer (GluN2B shown in *faded cyan*) **(D)** Dimer simulation (run 2), that captures probe binding at the GluN1-N2B interfacial cleft, with ifenprodil shown in magenta stick for reference (crystallographic binding site). We see that that the probe binding sites overlap some of the probe molecules seen in monomers (ACAM, pink as seen in N1 simulation, whereas the IPAM resemble those seen in GluN2B). Also the residues coordinating these probes are explicitly shown in stick representation.

Figure 43B shows the probes close (within 4.5 Å) to the UL interface (see also **Table 8**) and ifenprodil (green stick, Karakas et al et al., 2011), and the N1 dimer is shown in faded green for reference. This elucidates the intrinsic property of the interfacial region and polyamine coordinating regions to find binding partners. Similar sites are captured in run2, where we see ~ 25 hotspots near the UL interface, 75% of which are benzene molecules, shown in **Figure 44A**. In this run with benzene, we still capture the tendency of probe binding at the UL interface, but the clustering of benzene makes it harder to ascertain the actual number of hotspots that may bind to this region. **Figure 43B** shows probes from both run 1 and run 2. For run 1, we recover ~ 7 probes (3 IPRO, yellow and 4 IPAM, blue) that appear in the vicinity of the binding site. Interestingly, in run 2 we capture 7 probes, most of which overlap with probes from run 1 (circled regions, BENZ (grey) overlapping with IPRO (yellow), and IPAM (purple) overlaps with IPAM (blue) from run 2) and a common +ve charged probe (IPAM) binding site near the cationic N site piperidine group. This overlap shows the ability of the method to capture similar binding behavior in multiple runs even with different probe concentrations.

Overall the achievable binding affinity (from run 1) (we don't report for run 2 with benzene since some of the binding affinity may be driven by benzene's tendency to cluster) of the probes (6 probes that have overlapping counterparts, 3 IPRO, 3 IPAM) is ~54nM, which is ~10 fold higher than the experimental binding affinity reported for Ifenprodil (320nm) (Karakas et al et al., 2011). The abundance of +vely charged probes (3-4 IPAM) in the binding region

might contribute to this higher affinity, highlighting the preferential affinity of that region to bind +vely charged moieties. Thus, ligands that incorporate more +ve charged groups at the phenolic end of ifenprodil might offer higher binding affinities. The probes near the corresponding ligand binding region of the GluN1 monomer are shown in **Figure 43C**. We see that the hydrophobic end shows a cluster of IBUTs interacting mainly with F103. We don't see a distribution of +ve'ly charged probes near the phenolic end as seen for GluN2B, suggesting that interactions with GluN2B residues mainly drive the cluster of IPAM. A subtle balance between hydrophobic and +vely charged fragments might help in the identification of ligands with higher efficacy of binding at the GluN2-N1 interface.

5.4.2 Highlights from NMDA dimer simulations

Next, we explored the binding characteristics of the NMDA heterodimer that has been resolved in the presence of polyamines (3QEL, Karakas et al et al., 2011), and evaluate whether we can reproduce the interfacial ligand-binding site in this conformation. The first run was performed with the heterodimeric NMDA modeled on the polyamine bound structure (3QEL). At the end of a 60ns run we do not see any significant probe entry at the interface (**Figure 44B**), and the probes tend to cluster near the UL loops, and few at the GluN2B LL interface (similar to binding at the LL interface seen for GluN2B monomer simulations). This might be due to the ligand-free but closed starting conformation of the NMDA heterodimer. The accessibility to clamshell like motion facilitated by the bilobed structure of the NTDs may enable ligand binding, which is locked in the ligand-bound conformation. Hence, a more open conformation may enable us to observe probe binding near the cleft and subsequently near the interface.

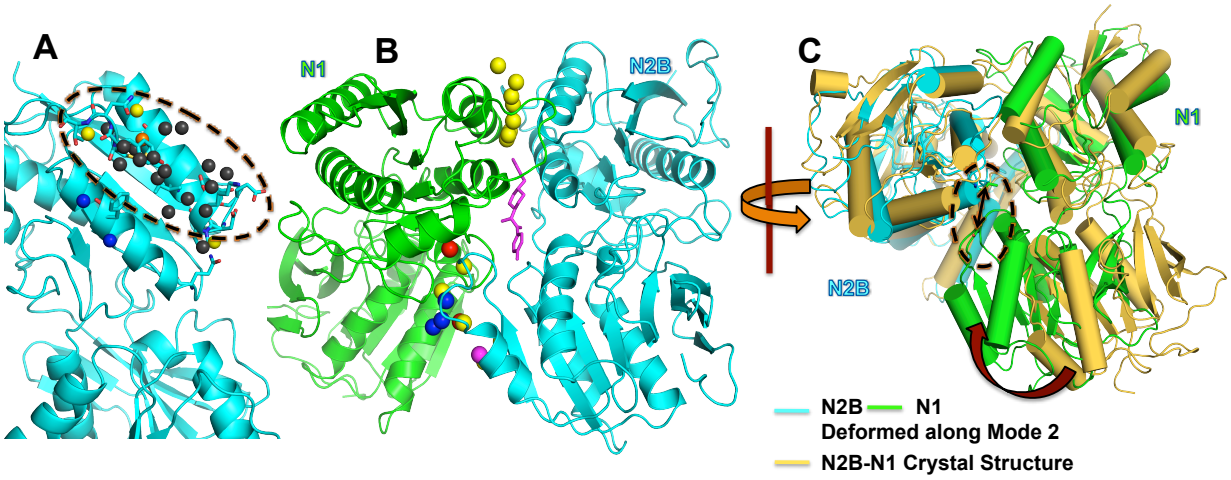


Figure 44. NMDA Monomer, Dimer probe binding and alternate conformation.

(A) Run2 (with 30% benzene) of GluN2B monomer and the probes within 4.5Å of UL interface residues (shown in stick). Of the 25 hotspots about 75% is benzene implying the clustering of benzene molecules (shown in black spheres) **(B)** Run 1 of NMDA heterodimer showing selected probes at the UL interface, and near the N1-N2B interface showing that these probes don't quite penetrate into the cleft to occupy the ifenprodil (in magenta stick) binding site. **(C)** Shows the superimposition of the NMDA heterodimer as it appears in the crystal structure (orange) and the conformation obtained after deformation along Mode 2 of ANM (See texr). The encircled region highlights the new dimer interface between GluN1 and GluN2B (double-headed arrow) as the GluN1 untwists in this mode (arrow).

Recently, hinge bending and twisting motions of the GluN1 subunit in the heterodimer has also been shown to modulate glutamate binding allosterically (Zhu et al., 2013) in NMDARs. Zhu et al., 2013 experimentally validated the accessibility of these motions in GluN1 that enable dimer interface rearrangements and an untwisting of the GluN1 conformation. This inherent flexibility of GluN1 and its characteristic motions were predicted from coarse-grained ENM analysis of proteins. They showed that the motions predicted by slowest or most global cooperative modes of ENMs for the NMDA heterodimer are accessible in the full-length receptor and may play a key role in influencing receptor-gating properties. The usefulness of

incorporating ENM predicted protein flexibility in docking studies has also been shown previously (Cavasotto et al et al., 2005).

Here, we generate an alternate conformation of the NMDAR heterodimer by deforming the NMDAR heterodimer along one of the slowest modes that Zhu et al., experimentally validated, namely Mode 8 or Mode 2 (non-zero modes). This primarily causes an untwisting in the GluN1 subunit leading to a conformation that forms a new GluN1-GluN2B interface such that distance between G200N1 and I329 N2B is ~8.4Å. Even though this doesn't lead to an opening near the ligand-binding cleft of GluN2B, there are subtle rearrangements in the UL of the GluN2B unit. We seek to understand the effect of this conformation on NMDAR ligand binding. **Figure 44**C shows the ANM deformed conformation of GluN2B (*cyan*) and GluN1 (*green*) overlaid on the ligand-bound closed state conformation (*pale yellow*), with the encircled region and arrows showing the large untwisting in GluN1 that enables a new inter subunit interaction.

Analysis of the probe binding shows the entry of probes in the GluN2B-N1 cleft that harbors the ifenprodil-binding site. We capture a probe distribution similar to that seen in monomer simulations, where the region otherwise coordinating the phenolic group of ifenprodil attracts positively charged isopropylamine molecules, while another IPAM mimics the interaction of the N+ piperidine with an additional acetamide binding site (*magenta* sphere in **Figure 43D**). Also, residues that coordinate the binding site shown in **Figure 43D** in stick representation closely resemble the crystal structure residues that coordinate polyamines (Burger et al., 2012; Karakas et al., 2011) in crystal structures and NMDAR models. However, the hydrophobic cluster seen in monomer simulations (IPRO and BENZ) is not captured here possibly due to the UL interface interactions between the monomers, and a subsequent lack of

monomer-probe interaction. Perhaps, longer simulations or alternate conformations that sample other conformations along ENM slow modes facilitating opening of the GluN2B unit and further rearrangements of the NTD-LBD UL interface, may enable us to recapitulate the ligand-binding site in its entirety in NMDAR dimers.

5.5 DRUGGABILITY OF DIMERIC AMPARS

Next, we proceeded to evaluate the "*druggabilty*" of AMPAR, GluA1-GluA3 (GluA4 is similar in conformation and flexibility to GluA2, and hence we expect a similar behavior as in GluA2). **Figure 45A-C** shows the probe distribution near the interfacial regions of the 3 dimers of AMPARs. In GluA3 dimer, there is a high affinity for probes to bind at the LL interface and also penetrate the UL interfacial region into the cleft. The high flexibility/relatively loose packing of the GluA3 LL and UL (Dutta et al., 2012; Sukumaran et al., 2011), Chapter 3) contributes to this affinity for probe molecules to bind at the interface. GluA1 is less flexible than GluA3, and thus fewer probes molecules can flood the interfacial region than that seen in GluA3, and the least flexible GluA2 with tight LL interface packing shows probes that cluster at the tips of the LL and UL, but cannot bind at the interface. Thus, the degree of LL flexibility mirrors the ability of the NTDs to bind probes in their interface. The main highlight from the simulations is the identification of a high-affinity ligand-binding site in the LL of GluA3 that we describe in details in the following section.

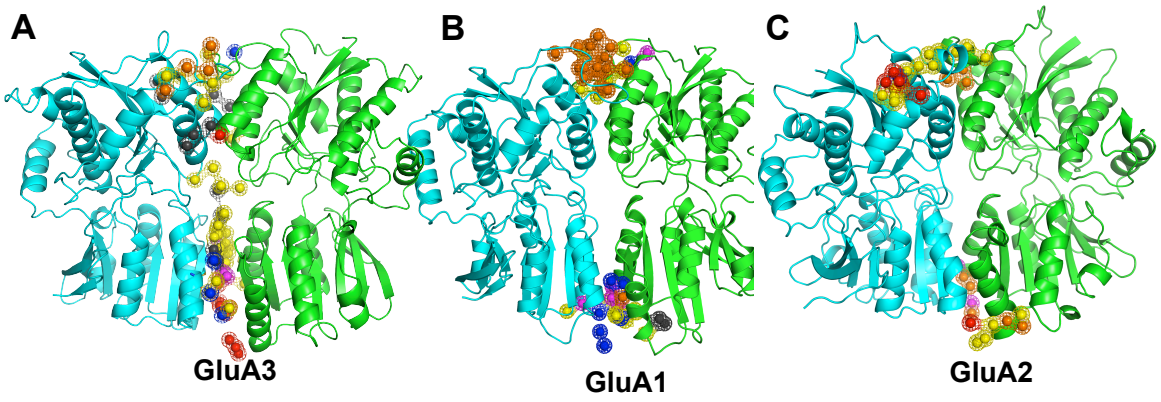


Figure 45. Simulation of AMPAR (GluA1-A3) dimers show distinct probe binding behavior consistent with intrinsic flexibility.

(A-C) COM of probes shown as spheres near the UL, LL interface of GluA3 (**A**), GluA1 (**B**) and GluA2 (**C**).

5.5.1 Pharmacophore features of GluA3 ligand-binding site

The LL interface of GluA3 is highly flexible. In simulations performed in the presence of water alone, this lobe tends to move apart and adopt a far more open conformation than that seen in crystal structure (Dutta et al., 2012; Sukumaran et al., 2011). In our current simulations, the binding of the probes in the LL interface stabilizes the tightly packed conformation and we do not see the same degree of LL opening as was previously observed. The probes that drive this behavior may be divided into two groups. **Figure 46** shows one cluster of primarily hydrophobic probes that bind in the pocket formed by M150. The probes interacting (within 4Å of M150) are highlighted here. The conformations of these probes are described by their average position over a few hundred -thousand frames in the course of the simulation. There are 8-9 probes that bind in close proximity of the M150 pocket that contribute mainly towards binding affinity (see **Table 9**). Additional probe molecules that do not directly interact with residues in the M150 binding pocket, but may contribute towards the shape and volume of the ligand as well as better

anchoring (for those that bind in deeper pockets) are shown in **Figure 46C**. The next step is to determine pharmacophore features (**Figure 46D**) that can be instrumental in virtual screening of compound libraries as a first pass for identifying ligands that conform to the shape, volume and specific chemical properties described by the model.

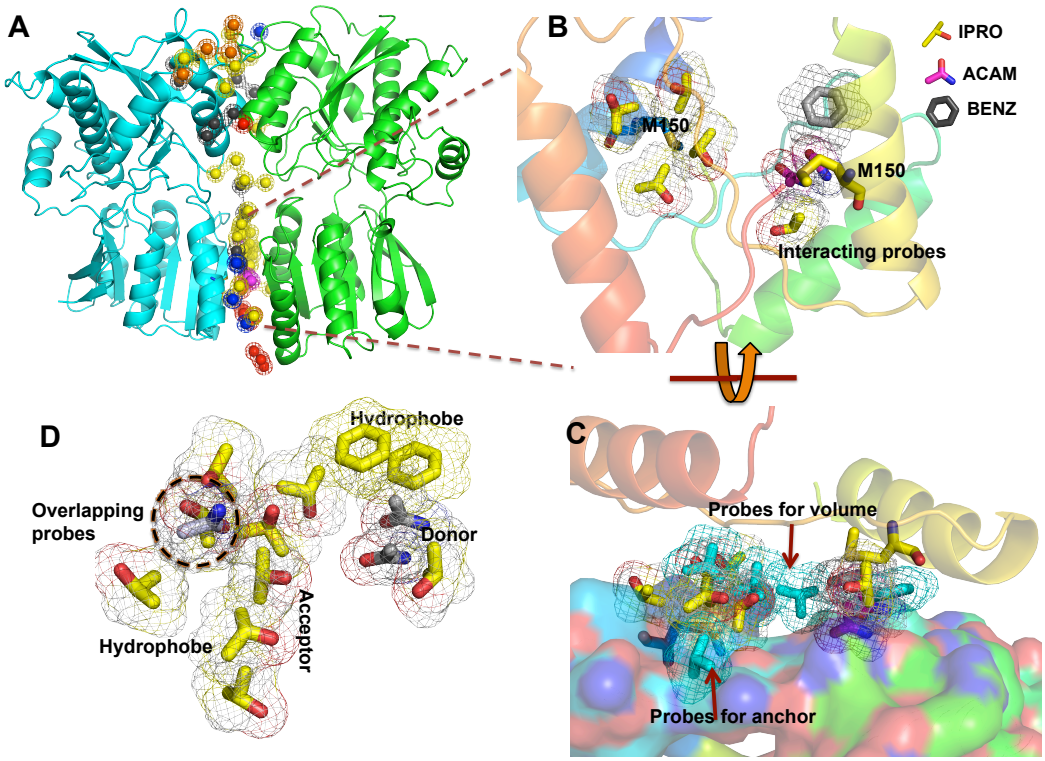


Figure 46. Pharmacophore feature of M150 binding pocket in GluA3 LL.

(A) GluA3 interfacial probes as shown in **Figure 45A**. **(B)** Shows a zoomed in view of the GluA3 LL probes that are primarily hydrophobic and bind in the pocket formed by M150. **(C)** Shows additional probes in (*cyan*) that contribute to shape/anchoring of the ligand-binding site. **(D)** Shows the pharmacophore features based on the probe compositions. Note that a single hot-spot may have contributions from multiple probes binding that region, and hence we may have overlapping functional groups (encircled). Hydrophobic groups are *yellow*, donors are *blue* and acceptors are *red*. Other residues that coordinate the probes apart from M150 include L146, Q147, V160, A162, and R163.

The Arg 163 and Arg 183 that are signature features of the AMPAR GluA3 as compared to its other counterparts (see **Figure 13** chapter 3) form a second pocket in the LL interface

where probes bind. The repulsion between these +vely charged residues contribute primarily to the destabilization of the GluA3 LL interface (Chapter 3). Hence ligands that modulate this pocket may have significant effects on the dynamics of GluA3, and thereby allosterically regulate channel action. **In Figure 47A-B**, we see the binding pocket formed by arginine residues and the surface representation of the GluA3 LL and probes that span this pocket. Note the abundance of acetates that coordinate the arginines, making it a primarily charged site, unlike the M150 hydrophobic binding region. Binding of acetates also prevents the same charge repulsion that was seen in "only-water" simulations that drove the LL's away from each other as seen in Dutta et al., 2012. The probes that characterize the pharmacophore features and those that contribute to the shape and volume of the pocket are highlighted in **Figure 47C**. The contributions of probes that bind the arginines are shown in **Table 9**.

Table 9. Druggability of GluA3 M150 and R163-184 LL pockets.

LL Probe Cluster	$\Delta G_{\text{probe binding}}$	Affinity	Probe contribution	charge
M150	**-13.08**	**0.34nM**	**6 IPRO, 1 ACAM, 1 BENZ**	**0e**
R163-R184	**-11.4**	**5.6nM**	**2 IPRO, 1 IPAM, 5 ACTT**	**4e**

Thus we find that there is a high potential for the identification of GluA3 specific molecules achieving nanomolar affinities that may modulate the function of GluA3 AMPARs.

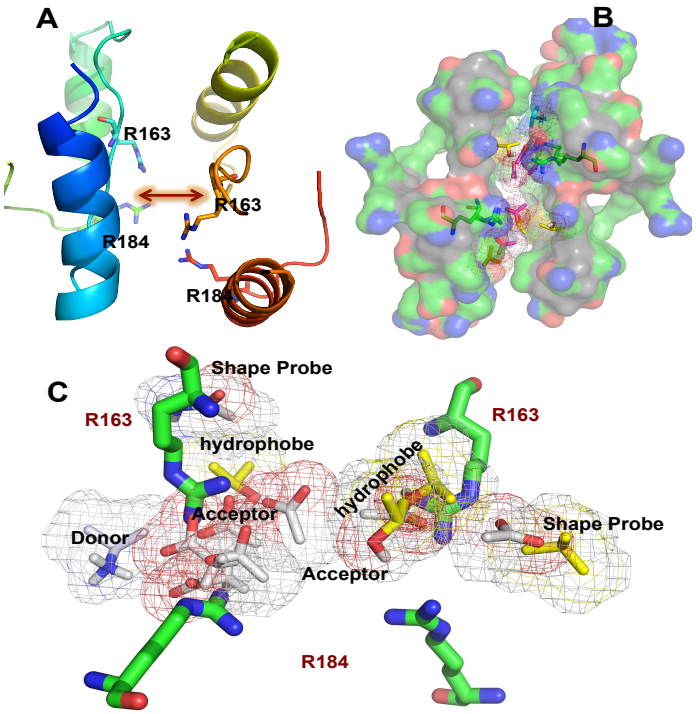


Figure 47. Pharmacophore model for Arg pocket in GluA3 LL.

(A) Shows the arginine pocket of GluA3 LLs. **(B)** Shows the surface representation of the lower lobes and the positioning of the probes that coordinate R163, R184, R183 and few other residues in this pocket (A153, V154, W158, Q159, E180, M181) **(C)** Similar to **Figure 46D**, where we formulate a pharmacophore model capturing the probe chemical properties, shape and volume.

5.6 CONCLUSION AND FUTURE WORK

The unexpected similarity in the intrinsic dynamics of AMPA- and NMDARs hints at the allosteric potential of non-NMDAR iGluRs and has propelled the evaluation of their "druggability". In this chapter, to achieve realistic detection of ligand-binding sites of iGluR NTDs and their maximal binding free energy, we performed MD simulations of our targets in the presence of drug-like fragments and water. First, we benchmarked this method by exploring the well-known ligand-binding landscape of GluA2 LBD. We found that binding of a single probe

molecule in the endogenous ligand binding site can drive domain closure akin to that seen during agonist binding. Subsequently, we explored the ligand-binding potential of all known iGluR NTDs in monomeric and dimeric forms. For NMDAR-GluN2B, binding sites for modulators like Zn^{2+} and phenylethanolamine compounds are known. Our method captures these binding sites with reasonable accuracy and also provides opportunities for designing compounds that may bind with better affinity than known compounds. Another striking result from an extensive analysis of all iGluR NTDs is the accuracy with which the probe-binding hot spots overlap with known dimer interfaces of monomeric NTDs. This opens new avenues whereby we can also accurately identify/predict protein-protein interfaces using this technique.

Simulations of AMPAR dimers provide a striking result, where we see extensive probe binding at the LL interface of GluA3 (not seen in GluA2 and GluA1 dimers), owing to a high degree of flexibility of GluA3. This suggests the presence of a high-affinity ligand-binding site in GluA3 LL interface. We characterize the pharmacophore feature of this site that is comprised of one hydrophobic pocket (near M150) and another –vely charged pocket (near R163, R184) based on the identity of the probes that bind in that region. Our model can serve as a first pass for virtual screen of compounds (from known libraries of drug-like compounds) to identify potential molecules that may conform to the pharmacophore features and shape of our model. The results can be further refined to identify top-ranking compounds that potentially bind with high efficacy selectively to GluA3, which can then be tested for functional significance. Thus this study opens a new arena for the identification of drug-like molecules that may bind to GluA3 and thus allosterically modulate ion-channel gating characteristics, widening the functional spectrum of AMPARs.

6.0 *EVOL*: BRIDGING BETWEEN PROTEIN STRUCTURAL DYNAMICS AND SEQUENCE EVOLUTION

Several recent studies have highlighted the significance of protein dynamics in accomplishing a wide range of biological functions, including protein-protein and -ligand interactions that often play a key role in enabling cell signaling, regulation and machinery (Bahar et al., 2010b; Bhabha et al., 2011; James et al., 2003; Marsh et al., 2012; Smock and Gierasch, 2009). Interestingly, recent studies have also observed that in many cases sequence variability goes hand in hand with structural dynamics (Glembo et al., 2012; Liu and Bahar, 2012; Marks et al., 2011; Micheletti, 2012; Worth et al., 2009; Zheng et al., 2005) and structural dynamics correlate with evolvability (Tokuriki and Tawfik, 2009). A recent systematic study demonstrated how sequence co-evolution plays an important role in enabling the adaptation to substrate binding while maintaining specificity at recognition sites that are usually characterized by enhanced mobility (Liu et al., 2010). It is now clear that a combined study of sequence evolution and co-evolution characteristics and structural dynamics can provide insights into evolutionarily selected functional mechanisms and their structural (dynamic) basis. Such studies would be particularly useful if they could be performed and visualized in an integrated computing environment, yet to our knowledge, existing software usually relate evolutionary properties to static structures (Ashkenazy et al., 2010; Eyal et al., 2007; Morgan et al., 2006; Wainreb et al., 2011), or they are exclusively dedicated to either sequence analyses (Waterhouse et al., 2009) or protein dynamics

(Eyal et al., 2006; Suhre and Sanejouand, 2004), and there is a need for a software package that could enable an efficient comparative study of sequence conservation and co-evolutionary patterns in the light of biomolecular structure and dynamics.

In this article, we propose to meet this goal by introducing a new release (v1.4) of our protein dynamics package *Prody* (Bakan et al., 2011) where functionalities of sequence analysis and database access have been added based on a method of approach delineated in a recent study (Liu and Bahar, 2012). A distinctive feature of *Prody* is its capability to extract mechanistic information from the principal component analysis (PCA) of ensembles of structures resolved to date for extensively studied proteins (e.g. drug targets) (Bakan and Bahar, 2009). In the new release, the structural ensembles are analyzed together with the data extracted from multiple sequence alignments (MSAs) (a.k.a. ensembles of sequences) for protein family members, in order to provide a comprehensive view of the family dynamics in the light of evolutionary propensities. We have added several new modules and command line *applications* named **'evol'** to calculate sequence conservation and co-evolution using information-theoretic approaches (Cover and Thomas, 2001). Most of the new modules are written in *C* for high performance. Based on Python, the application programming interface (API) enables all the functions to be easily integrated into the user's own code. To our knowledge, this is the only package that integrates protein dynamics and sequence conservation and coevolution data.

A large number of methods exist for identification of coevolving residue pairs; those based on local pairwise coupling (Atchley et al., 2000; Fodor and Aldrich, 2004; Kass and Horovitz, 2002; Lockless and Ranganathan, 1999; Martin et al., 2005) from MSAs, and more recently those based on inference of global statistical models (Burger and van Nimwegen, 2010; Jones et al., 2012; Morcos et al., 2011; Weigt et al., 2009) from MSAs. In this study, apart from

an illustration of the use of **'*evol*'** we also probe deeper into local *vs* global methods for identifying co-evolving pairs and their ability to extract information about 3-D structural contacts. Most coevolution methods uncover relationship between pairs that are structurally distant and thus have low signal to noise ratio. This has been ascribed to the following reasons i) background noise arising out of phylogenetic relationships between sequences in the MSA (Fodor and Aldrich, 2004; Noivirt et al., 2005; Wollenberg and Atchley, 2000) ii) low and biased sequence sampling and small sequence diversity (Cline et al., 2002) and iii) indirect transitive relations propagating through chains of correlated residues (Morcos et al., 2011; Procaccini et al., 2011; Weigt et al., 2009). The global methods of sequence correlation take advantage of currently available large sequence samples (like the Pfam database) and infer direct relations (removing indirect associations) and have been used for predicting protein structures due to their ability to predict precise structural contacts (Hopf et al., 2012; Marks et al., 2011; Marks et al., 2012). On the other hand, local methods like mutual information with a correction for phylogenetic noise have also been seen to gain an increase in signal-noise, even though they have not been directly used for structure prediction (Dunn et al., 2008).

Here, we illustrate the utility of our recently developed **'*evol*'** module by performing a hierarchical study of the relationship between conservation and coevolution, inferred from both local and global methods, for the iGluRs, NTD and the homologous mGluR LBD. We use a bottom up approach, where we start with analyzing a subfamily of iGluR, followed by addition of larger subfamilies to identify various levels of structural constraints (inferred from sequence patterns). Also, we present a key property of evolutionary dynamics, where residues that are conserved in sub-families tend to become coevolving when considering larger subfamilies, suggesting their importance in function. Coevolving residues also form a network of connections

that highlight possible channels of information propagation necessary for allosteric signaling in iGluR NTDs.

6.1 NEW APPROACH: DESCRIPTION AND FUNCTIONALITY

Our recent investigation of a representative dataset of enzymes revealed significant relationships between evolutionary properties and structure-encoded intrinsic dynamics of proteins (Liu and Bahar, 2012). The basic approach therein was to perform a comparative analysis of two sets of properties, based on sequence variations and structural dynamics, respectively: (a) sequence conservation profile and coevolution propensities between residues, deduced from the statistical analysis of the MSA of the protein family (after refinement) obtained from the Pfam database (Punta et al., 2012) and (b) residue mobility and cross-correlations (covariance) between residue fluctuations in 3D space, derived from the GNM analysis (Bahar et al., 1997; Bahar et al., 2010b) of protein dynamics. The latter is performed for a representative member of the protein family, which has been structurally resolved and deposited in the Protein Data Bank (PDB) (Berman et al., 2000). Adopting a similar premise as in the earlier work, the extension of *ProDy* is designed for fast and efficient comparisons between evolutionary patterns and structural dynamics applicable to any protein family. In the following sections we first describe the analysis steps and API features, and then illustrate `***evol***` applications. Next, we provide a detailed analysis of sequence-structure relationships by studying the hierarchical constraints in the NTD/LBD of the glutamate receptor (GluR) family.

6.1.1 Input for Sequence Analysis in *ProDy*

The new *database* module of *ProDy* offers functions that enable users to search the Pfam database, and retrieve the Pfam accession number or ID for each protein family. The input to the *search* function can be a UniProt ID, a PDB ID or a protein sequence, in which case a sequence search is initiated using Pfams' search module, and the resulting matches are retrieved. Note that one protein sequence may have multiple hits across different Pfam families (>13,000 manually curated families as of release v26.0 (Punta et al., 2012)), and it is up to the user to analyze any particular family. The *e-values* associated with each match are displayed to provide an assessment of the significance of matches. The Pfam ID obtained from the search form the input to the *fetch* function that downloads the MSA for the query family in one of the formats supported by Pfam.

6.1.2 MSA IO and Refinement

ProDy implements classes and functions that can parse the MSA files retrieved from the Pfam database. Notably, parsing uncompressed MSA files using underlying *C* modules greatly improves I/O efficiency. Once we parse the MSA, these classes provide efficient ways of sampling from the MSA in a user-defined way. We can filter the MSA based on the labels associated with each sequence (mainly their UniProt IDs) such that only selected categories of sequences would be retained (e.g., human sequences). Most importantly we can slice the MSA, i.e modify either the sequences (rows) or the positions (columns) of the MSA, to retain specific regions or sequences. This is important in our analysis where the first refinement step involves removal or slicing of the columns in the MSA such that the gaps for a given sequence

(corresponding to the representative PDB structure) can be eliminated. This step is necessary because we focus on the portions of the MSA that have structurally resolved counterparts in the PDB. Further refinement includes removal of sequences that have more than 20% gaps, as well as those above a user-defined sequence identity threshold so as to obtain statistically significant results upon further analysis.

6.1.3 Shannon Entropy for Conservation

Functions in *ProDy* allow for the calculation and visualization (see **Figure 48A**) of Shannon entropy, $S_i = - \mathbf{\Sigma}_{xi} [P(x_i) \ln P(x_i)]$ for each sequence position (column) i of an MSA. Here $P(x_i)$ designates the probability of amino acid type x at position i, and the summation is performed over all 20 types of amino acids. Shannon entropy provides a measure of variability or tolerance to mutation (Cover and Thomas, 2001). It varies in the range $0 \leq S_i < \ln 20 = 3.00$, the lower and upper corresponding to fully conserved positions and random/uniform distribution of amino acid types at position i, respectively. The calculations also take into account ambiguous amino acids (like Asx (Asp or Asn), and modified amino acids like selenocysteine, pyrrolysine) or gaps. *ProDy* permits users to view conservation profiles (as a function of sequence index) with the help of color-coded ribbon diagrams, which may be readily compared to those generated based on conformational mobility profiles of amino acids along the sequence (**Figure 48B**).

6.1.4 Consurf for Conservation

Consurf is a well known web server that calculates the conservation profile of residues in a protein sequence identifying regions that are important for function and structure. The evolutionary rate calculation is based on emprical bayesian or maximum likelihood approach that captures the underlying stochastics associated with evolutinary processes better (Ashkenazy et al., 2010; Landau et al., 2005). The MSAs were uploaded to the Consurf server along with the associated structure and calculations were run with default settings at http://consurf.tau.ac.il/.

6.1.5 Mutual Information for Residue Coevolution

Mutual information (MI) is a metric for assessing the coevolution propensity between two positions (columns) in a given MSA, thus providing insights into structurally or functionally important sites whose mutation tend to be accompanied by compensating (or correlated) substitutions at other sites (Cover and Thomas, 2001). In a strict sense, it measures how knowledge of amino acid distribution at the sequence position *i* of the MSA reduces the uncertainty in the distribution of amino acids at position *j* (see (Atchley et al., 2000; Liu and Bahar, 2012; Martin et al., 2005). High MI values indicate the tendency of the corresponding residue pairs (or sequence positions) to coevolve, or undergo correlated substitutions. *ProDy* supports the calculation and visualization of mutual information based on the classical definition

$$\mathrm{I}(i,j) = \sum_{x_i=1}^{21} \sum_{y_j=1}^{21} P(x_i, y_j) \log \left[P(x_i, y_j) / P(x_i) P(y_j) \right] \qquad (20)$$

as well as its normalized variant $\mathrm{I_N}$ (Martin et al., 2005) or corrected variant $\mathrm{I_P}$ (Dunn et al., 2008). The latter is defined as (Dunn et al., 2008)

$$MI_P(i, j) = I(i, j) - APC\ (i, j) \quad (21)$$

where $P(x_i, y_j)$ is the joint probability of occurrence of amino acid types x_i and y_j at the respective i^{th} and j^{th} positions, and APC(i, j) is the average product correction, $APC(i,j) = \langle I(i)\rangle\langle I(j)\rangle / \langle I\rangle$, where $\langle I(i)\rangle$ is the average of I(i, j) over all sequence positions j, and <I> is the average over all i and j. Signals obtained from MI calculations suffer from noise resulting from sample size, phylogenetic effects and entropic effects (Atchley et al., 2000; Martin et al., 2005; Tillier and Lui, 2003), hence the introduction of normalized or corrected variants.

The ***evol*** application of *ProDy* permits to view the mutual information for all pairs of amino acids with the help of heatmaps, the axes of which refer to sequence positions along the query protein. MI can be calculated very fast if large memory allocation is possible (turbo mode), or a slower more memory efficient implementation is adopted. *ProDy* also enables rank ordering of MI scores or MI values post z-score normalization (Martin et al., 2005) to identify top pairs that tend to coevolve. We can apply certain filters requiring, for example, that pairs should have sequence separation larger than three positions, or C^a distance cutoff of 15 Å to identify long-range (along sequence or space) correlations (see below).

6.1.6 Direct Information for Residue Coevolution

Covariance techniques like mutual information infer statistical correlation from constraints seen in amino acid distribution patterns in MSA. In simple covariance methods, apart from phylogenetic and entropic noise (corrected for as seen above), a large percentage of confounding correlations occur due to indirect associations. For example, if A→B (A coevolves with B), B→C, we often see correlations like A→C, even though they are not in close proximity in tertiary structure. Direct Coupling Analysis (DCA) is a global method for inferring residue-pair

correlations that "disentangles" direct interactions from indirect ones to a large extent (Morcos et al., 2011; Procaccini et al., 2011; Weigt et al., 2009). DCA has been shown to have higher accuracy in predicting true contacts and hence has been successfully used to facilitate prediction of 3D structure in soluble and membrane proteins (Hopf et al., 2012; Marks et al., 2011). Certain other methods that infer global characteristics include a Bayesian approach (Burger and van Nimwegen, 2010), and a sparse-inverse matrix factorization technique (Jones et al., 2012; Procaccini et al., 2011), which also show better predictive ability.

The main premise of the method is to infer a global statistical model for the protein sequence $P(A_1, A_2, \ldots A_N)$, such that the marginal distribution up to one and two positions is coherent with observed data, or $P_i(A_i) = S\ P(A_1, A_2, \ldots A_N)$, for all $A_k \neq i$, $\cong f_i(A_i)$, and $P_{ij}(A_i, A_j) = S\ P(A_1, A_2, \ldots A_N)$, for all $A_k \neq i,j$, $\cong f_{ij}(A_i, A_j)$. Whereas, a large number of models can satisfy these constraints, adopting the maximum entropy principle ensures a mathematical form for $P(A_1, A_2, \ldots A_N)$ given by

$$P(A_1, A_2, \ldots A_N) = \frac{1}{Z} \exp\{-\textstyle\sum_{i<j} e_{ij}(A_i, A_j) + \sum_i h_i(A_i)\} \qquad (22)$$

where, $e_{ij}(A_i, A_j)$ is pairwise coupling and can be estimated by a heuristic message passing approach (Weigt et al. 2009), or by mean-field approximation (Morcos et al. 2011) that yields $C_{ij}(A_i, A_j) = f_{ij}(A_i, A_j) - f_i(A_i)\, f_j(A_j) = e^{-1}(A_i, A_j)$. Following this, a quantity called direct information (DI), similar to MI in formulation, is introduced to rank pairwise couplings,

$$DI(i,j) = \sum_{i=1}^{21}\sum_{j=1}^{21} P^{dir}(A_i, A_j) \log\left[P^{dir}(A_i, A_j) / f(A_i) f(A_j)\right], \qquad (23)$$

where $P_{ij}^{dir}(A_i, A_j) = \frac{1}{Z_{ij}} \exp\{\sum_{i<j} e_{ij}(A_i, A_j) + \sum_i \tilde{h_i}(A_i) + \sum_j \tilde{h_j}(A_j)\}$, and $\tilde{h_i}$ are local "fields" estimated from adherence to single residue counts, and $P_{ij}^{dir}(A_i, A_j)$ is the global estimation of pair probabilities consistent over all i,j in the MSA (Morcos et al. 2011, Marks et al. 2011).

Another important feature associated with this method as implemented by Morcos et al., 2009 is the reweighting of MSA sequences (M_{eff}) for removing bias associated with sequence sampling (high number of similar sequences) (See reference for details). Also global inference requires large MSAs size (~ 1000 M_{eff}) for accurate statistical sampling, which has been made possible recently due to larger number of genomic sequences obtained through fast sequencing techniques. The "*evol*" application does not incorporate this method yet, but in future may be modified to implement such global statistical methods for coevolution analysis.

6.1.7 Analysis of Structural Dynamics

ProDy (Bakan et al., 2011) is designed for inferring structural dynamics from PCA of experimental structural datasets or snapshots from molecular dynamics simulations, as well as predictions from theoretical models and methods including GNM (Bahar et al., 1997) and the anisotropic network model (ANM) (Atilgan et al., 2001). The current version 1.4, now enables comparison between sequence evolutionary patterns and structural dynamics of protein families, as demonstrated by Liu and Bahar using their GNM analysis. Flexible features of *ProDy* allow for comparison of the mobility profiles associated with a few dominant (low frequency) modes as well as the fluctuations originating from all modes of motions as predicted by GNM/ANM analyses or inferred from PCA analysis of experimental structural datasets (**Figure 48A**).

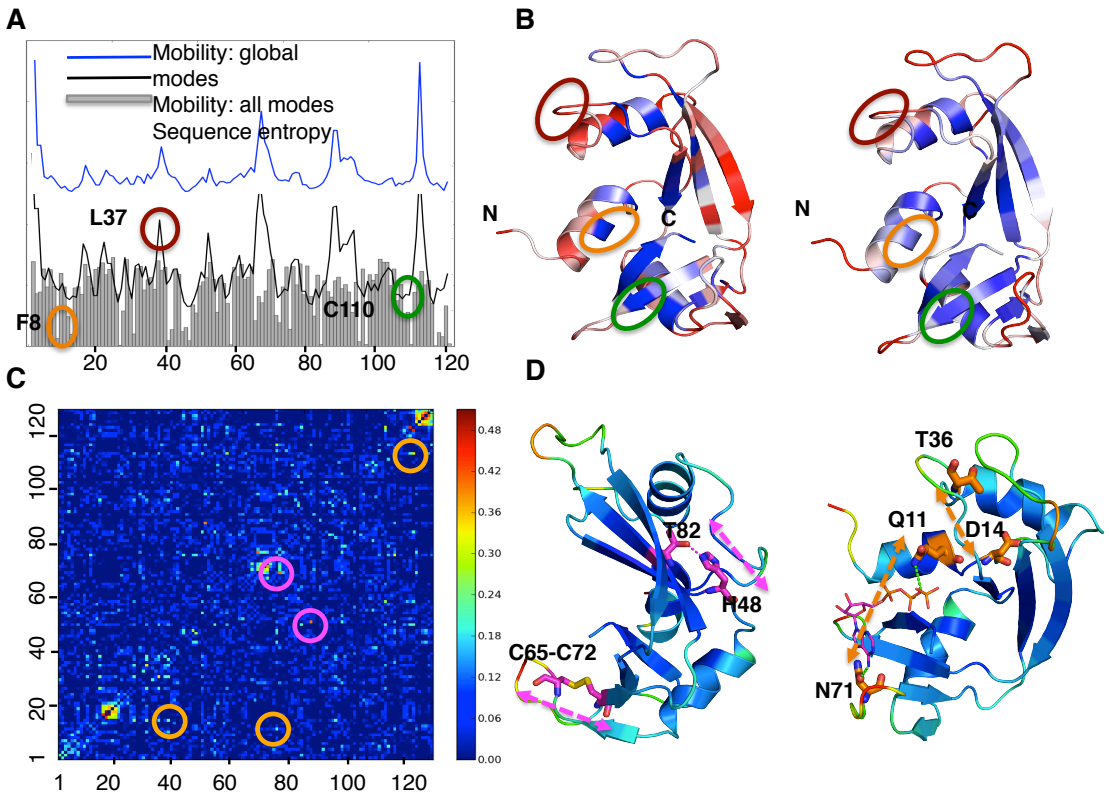


Figure 48. Comparative analysis between sequence conservation, coevolution patterns and global dynamics.

The above exampleis for a member of the RNase A family (PDB ID:2W5I –*B*) **(A)** Correlation between sequence entropy (gray bars) and mobility profile based on all modes (*black* curve), and global modes (combining eight lowest frequencymodes, *blue* curve) predicted by the GNM **(B)** Comparison of the color-coded ribbon diagrams generated using residue conservation (*left*) and intrinsic conformational mobility (*right*). Highly conserved (low entropy) residues, colored *blue*, also have lower mobility (*blue*, *right*). Conversely, highly variable residues (high entropy;*red*) tend to occupy highly mobile regions (*red*, *right*). A few residues are highlighted (encircled in A and B) to ease the comparison. **(C, D)**: Mutual information (MIp(*i, j*)) for each residue pair, and pairs distinguished by their high coevolution tendencies. The bright points (*cyan* to *red*) in the heat map refer to pairs that have high coevolution propensities. A number of sequentially distant pairs (≥ 6 intervening residues) are highlighted by circles, including spatially close (*magenta*) or distant (*orange*) pairs shown in panel D. Notably, (C65, C72) forms a disulfide bridge, and (T82, H48) make side chain (polar) interactions (*left* diagram). And the pairs (N71, Q11) and (T36, D14) are presumably involved in allosteric interactions (*right* diagram). The *right* diagram in panel D displays the RNase A crystallized in the presence of an inhibitor-like substrate (*thin stick* representation) (Holloway et al., 2009). Q11 and N71 form hydrogen bonds with the substrate possibly to maintain binding/recognition specificity, while D14 (near the binding site) shows long-range coevolution with a distant part of the protein (T36) indicating an allosteric communication pathway.

6.2 ILLUSTRATION OF EVOL APPLICATIONS

In addition to the API, we have developed a set of *applications* assembled under the '**evol'** command, which work similar to those under the '**prody**' command, e.g. `**evol search**`. These applications provide a simple yet efficient way of examining and visualizing the evolutionary and dynamic signatures of protein families.

Figure 48 illustrates the results from comparative examination of sequence conservation and structural dynamics features an RNase A family member (PDB ID: 2W5I) (Holloway et al., 2009). A search in Pfam using `**evol search 2w5i**` command yielded the Pfam family id PF00074, "RnaseA" corresponding to PDB 2W5I. We downloaded the full MSA from Pfam (`**evol fetch RnaseA**`) and removed all gapped columns in the MSA for sequence with label "RNAS1_BOVIN" (`**evol refine RnaseA_full.slx –l RNAS1_BOVIN --seqid 0.98 --rowocc 0.8**`) (UniProt ID corresponding to PDB ID: 2W5I), resulting in 121 positions corresponding to residues 1-121 in the PDB. Following removal of sequences with > 20% gaps (minimal row occupancy of 0.8), and the sequences that share 98% identity, the MSA was refined to contain 383 sequences with 121 residues.

Figure 48A shows the mobility profile obtained from GNM for RNase A (using the PDB structure 2W5I). The *black* curve shows residue mobility driven by all modes, and the *blue,* that by the global modes (based on eight lowest frequency GNM modes) that cumulatively account for 29% of the structural dynamics. The gray bars show the Shannon entropies or conservation profile evaluated using the refined MSA and *conserv* application (`**evol conserv RnaseA_full_refined.slx**`. The ribbon diagrams in **Figure 48B** show the protein color-coded by its conservation profile (*left*) and mobility (*right*; all modes). The correlation between conservation and mobilities is apparent in **Figure 48B**: residues that have lower mobilities (hinges, catalytic

residues) are conserved (*blue* regions). **Figure 48C** shows the coevolution propensities based on mutual information (I_p) values, obtained with the *coevol* application (`**evol coevol RnaseA_full_refined.slx -S -c apc**`). *ProDy* provides the rank-ordered lists of the co-evolving pairs based on their mutual information (MI, MI_p or MI_N) values. The application *rankorder* creates a file that lists the top-ranking coevolving pairs. The list contains residue pairs separated by a preset number (e.g. `**evol rankorder RnaseA_full_refined_mutinfo_corr_apc.txt -p 2W5I_1-121.pdb –-seq-sep 3** ` for 3 intervening residues). Alternatively, spatially distant pairs above a certain threshold distance (e.g. those separated by at least 15Å based on their C^{α}-C^{α} distance may be extracted (`**evol rankorder RnaseA_full_refined_mutinfo_corr_apc.txt -p 2W5I_1-121.pdb --use-dist --min-dist 15**`). The two diagrams in **Figure 48D** highlight residue pairs that were distinguished by their high evolutionary propensities (encircled in panel **C**) after the two respective types of filters. The *left* diagram in panel **D** highlights two pairs (*magenta*) that exhibited strong signals, presumably due to the critical importance of the disulfide bridge (between C65 and C72), or the tertiary contact between T82 and H48. Other (not shown) pairs distinguished by their top-ranking co-evolutionary propensities, which make close tertiary contacts, include (Y25, D14) and (A109, N71). Panel **D** *right* diagram displays residue pairs that exhibit high co-evolution propensities, despite being spatially distant (> 15Å). Notably, the pair (Q11, N71) coordinates the substrate, while the (T36, D14) pair establishes the communication between the substrate-binding site and the distal helix 2.

The installation and use of *ProDy* 1.4 with `**evol**` extension requires Python 2.7 or higher and Numpy (scientific computing tool for Python) version 1.4 or higher. For sequence analysis, MSAs of the protein families are obtained from Pfam (Punta et al., 2012), and for protein dynamics, structure data are obtained from the PDB (Berman et al., 2000). Visualization of

results is accomplished by Python plotting library Matplotlib (Hunter, 2007). Detailed descriptions of how to use the different aspects of *evol* can also be in the tutorials section of *Prody-evol* available at http://www.csb.pitt.edu/prody/tutorials/evol_tutorial/index.html.

6.3 HIERARCHICAL ANALYSIS OF THE GLUR NTDS

The three pharmacologically distinct classes of iGluRs, named and categorized based on their efficacy towards binding of AMPA, NMDA and Kainate. These different subfamilies encompassing the iGluR family have been described in Section 1.1, Figure1.

6.3.1 Sequence Alignment and Refinement Steps

Figure 49 illustrates, the hierarchical view of the MSAs used for evolutionary analysis, where we compare the distinct subsets of iGluR NTD, starting with AMPA (*A*), followed by AMPA, Kainate (*AK*) (that are structurally and sequentially higher in similarity), then encompassing AMPA, Kainate and NMDA (*AKN*). We also study the effect of addition of homologous mGluR LBD sequences (*AKNM*), followed by analysis of the entire Pfam ANF receptor family (*ALL*) encompassing eukaryotic and prokaryotic members (Pfam ID: PF01094). The steps taken for analysis are shown in **Figure 50**. The AMPA, Kainate, NMDAR and mGluR contain hand curated set of sequences that are aligned using *hmmbuild* and *hmmalign* packages from HMMER3 (Finn et al., 2011). Each individual set (*AK, AKN, AKNM*) is built from unaligned sequences of distinct subsets, eg: *A*, *K* unaligned and then merged to obtained *AK* unaligned,

followed by alignment using HMMER. This is followed by a rigorous refinement step, as shown in step 2 of **Figure 50**, using '*refineMSA*' API module of `***evol'*** as described in Section 6.2.

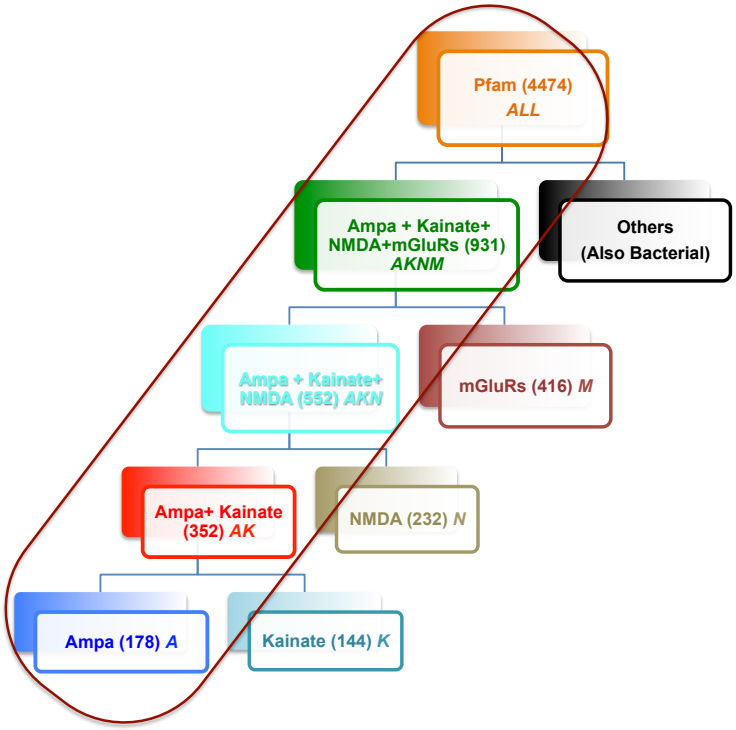


Figure 49. A hierarchical analysis of GluR family sequences.

Shows the subfamilies used for MSA based hierarchical analysis, starting at the lower level with AMPA family, with the number is parenthesis indicating the total number of sequences following refinement and the abbreviation (*A*) in italics. The next level is the addition of Kainate sequences (*K*), followed by NMDA(*N*) to AMPA and Kainate, and then by mGluR sequences (*M*). The top level alignment is obtained from Pfam (PF01094).

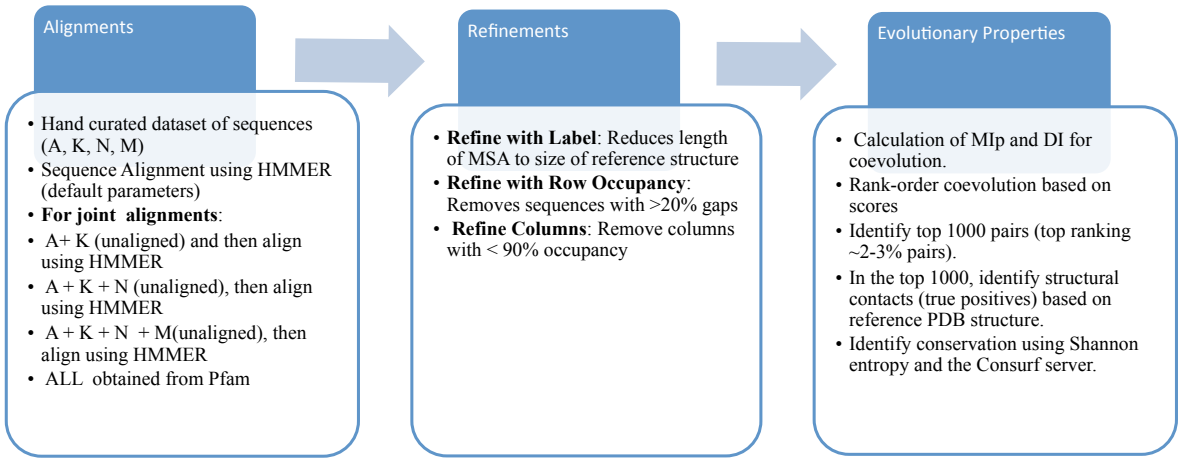


Figure 50. Flowchart of steps involved in analysis.

The first step in refinement, w.r.t to label, enables mapping onto the PDB structure. Label refinements for all subsets were done based on Uniprot ID, "*GRIA2_RAT*" that corresponds to GluA2 structure resolved in PDB with *ID* 3HSY (Sukumaran et al., 2011). The next step maintains a good quality of alignment, necessary for noise reduction. Note that the last step in refinement, where we remove columns with > 10% gaps, is mostly applicable to MSAs with smaller no. of sequences (*A, AK, AKN*). In a small set, a large amount of signal originates from columns that have >10% gaps, since presence of gaps (considered as 21st amino acid) in similar positions indicate correlation. This further highlights the extreme dependence of coevolution signals on the MSA, its size and quality. In order to be consistent in our analysis, we have maintained the column refinement criteria even for larger sequence sets (*AKNM* and *ALL*). The numbers in **Figure 49** in parenthesis indicate the total number of sequences that are present in each set following all refinement steps.

6.3.2 Conservation Analysis

Next, we calculate conservation scores obtained from the Consurf server and identify highly conserved residues as those that get a Consurf score of 9. Whereas Shannon entropy is a good measure of sequence diversity; here we wanted to use a more robust method that takes evolutionary relationships into account. **Figure 51A** shows the number of residues identified as being highly conserved for each of the sequence sets. As expected the total number of conserved residues decreases as we add more family members. The decrease is starker going from AK to AKN (from 40 to 15) and does not decrease much further in larger sets. The NMDARs are sequentially quite varied than the AMPAR's and Kainate's (though they preserve structural similarity) and addition of these sequences results in more diverse sequences in the MSA, resulting in decrease in conserved residues. It is interesting to note how conservation patterns change in subsequent subsets and those that are maintained, providing insights into functional constraints. **Figure 51** shows the residues that are conserved across different groups eg: 2 residues are conserved in all sets, 6 in A, AK, AK and AKNM and so on. **Figure 51C**, shows the residues mapped on the structure of GluA2.

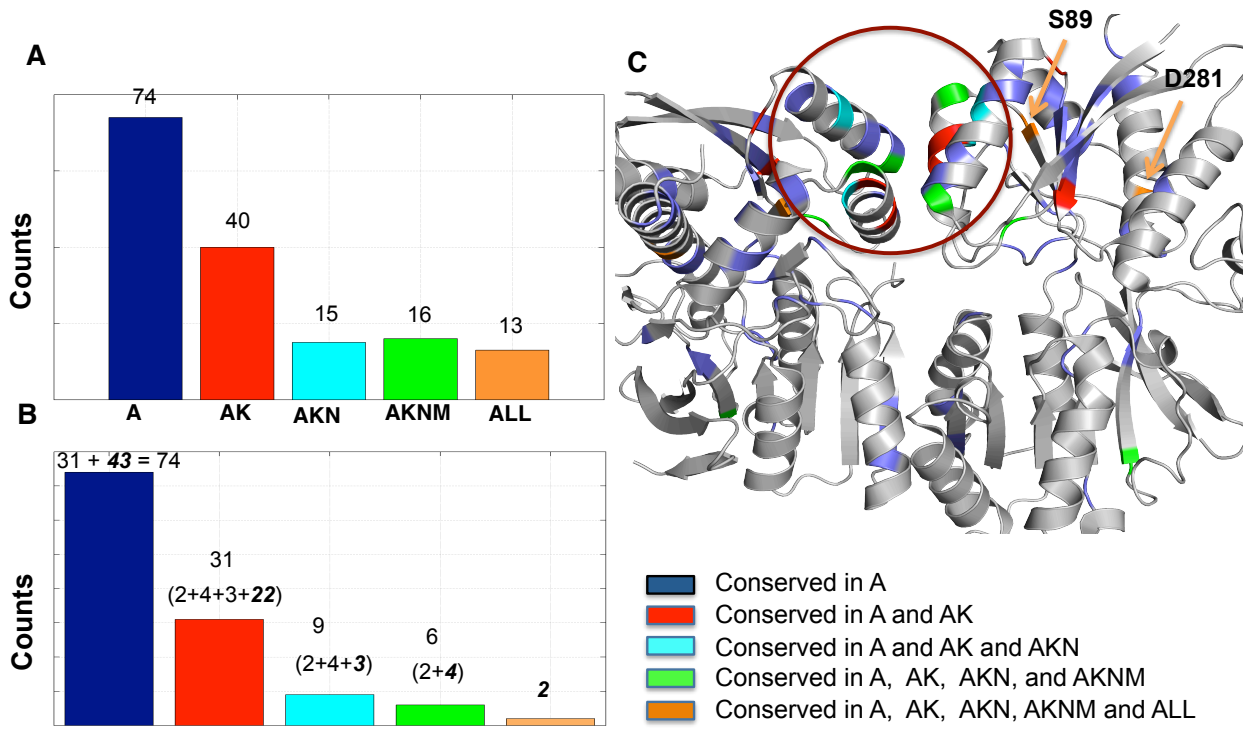


Figure 51. Conservation profile across all subsets of sequences.

(A) Shows the counts of the number of highly conserved residues (those obtaining a score of 9 from Consurf server) in each of the alignments subsets (*A – ALL*). **(B)** Shows the number of conserved residues that are common in different subsets. Note that the blue bar in A,B are the same. The *orange* (last) bar in B, shows the number of conserved residues that are common in all subsets (intersection of all residues *in A, AK, AKN, AKNM* and *ALL*). The green bar shows the residues conserved in A, AK, AKN, AKNM. The total number is 6, and the numbers in parenthesis show that 4 (in bold, italics) are unique to this set, and 2 are from the previous set (if it is conserved in all groups, then it would be conserved in the intersection of a smaller group). Similarly for conservation in A, AK, AKN (*cyan*) groups where 3 (bold, italic) is unique to this set, and 2, 4 are from previous sets, making a total of 9, and so on. These residues from B, are mapped onto the structure in C (PDB ID: 3HSY), with the two residues in the orange bar explicitly marked. The encircled region highlights the interfacial residues that are highly conserved in smaller sets and gradually become less conserved in higher sets.

The two highlights evident from **Figure 51C** are i) There are larger numbers of conserved residues in the UL than the LL of NTDs. This higher variability in sequence is also reflected in structural diversity in the packing of LL, range of flexibility and potential role in

signaling during allosteric modulation (Dutta et al., 2012; Sukumaran et al., 2011). ii) The UL interface is most conserved in AMPARs (*blue*), and some important conserved residues are maintained across AK (*red*) and AKN (*cyan*) (S49, C57, S81) and also to a certain extent after adding mGluR (in AKNM) (green). However, these disappear when considering the entire Pfam family (*orange*). The two most conserved entities that survive are S89 and D281 that are central to maintaining the structure of the NTDs. The UL interface has been associated with initialization of oligomer association in dimeric GluRs. Hence, we see stronger signals when analyzing these sub-families. On addition of prokaryotic homologs, we see that this strong conservation is no longer maintained especially since a larger number of these homologues are not dimeric but function as monomers. Thus, we capture subtle functional associations on analysis of conservation patterns from our hierarchical analysis.

6.3.3 Coevolution Analysis

Next, we focus on evaluating the performance on the prediction of structural contacts from a) mutual information with correction for phylogenetic background noise (eq: 21, MIp) and b) direct information (eq: 23, DI) with correction for transitive relations. **Figure 52** shows the number of true positives (TPs are residues with sequence separation > 5 and C^{α} distance < 10Å), predicted out of the top 1000 coevolving pair predictions. Notably, the number of TPs increase for both MIp and DI as we increase the number of sequences (more evident for MIp than DI). This could be due to the fact that there are higher sequence diversity enables us to better capture the evolutionary constraints whereas smaller subsets have more conserved residues that by definition are not coevolving. Also, other noise factors are compounded in smaller subsets that result in correlation signal (between distant pairs) almost equal in magnitude to those in TPs

(evident from almost uniform distribution of TPs over top 1000 predictions, **Figure 54**). Even DI (which reduces indirect associations) captures a large number of indirect associations and noise, possibly due to smaller number and diversity in sequences, which does not provide enough data to formulate an accurate statistical model. This changes when we look at the database of the entire pfam family, where both methods do much better in predicting TPs, with DI outperforming MIp. Thus protein families that incorporate a large amount of sequence data capture the evolutionary dynamics better and hence are more suitable for prediction of tertiary contacts necessary for structure prediction.

Next, when we identify the intersections or common predicted TPs that come from MIp and DI (diagonal of Table 1, and *yellow* bar in **Figure 52A**). Remarkably, in smaller subsets there is little overlap between the predictions in smaller sequence sets (4 for AK, 5 for AKN and so on). It is unclear why the contacts predicted by the two methods are so diverse and further analysis needs to be performed to understand the nuances. It is possible that certain direct and independent associations are being filtered out by DI (present in MIp) (while trying to capture dependencies associated with the entire MSA), whereas, noise due to entropic and chain-like transitive associations in MIp hide signals that are captured by DI. Nevertheless, the signals that are retained in both methods (especially for A, AK, AKN and AKNM) are the strongest signals and are shown in explicit cartoon representation in **Figure 52B-C** in colors. Again the upper lobe interfacial region has strong correlations with structural elements in the UL suggesting their importance in maintaining structure and function. Also, residues in the cleft and those in the LLs are shown that may be important for structure and propagation of signals further downstream of the NTD.

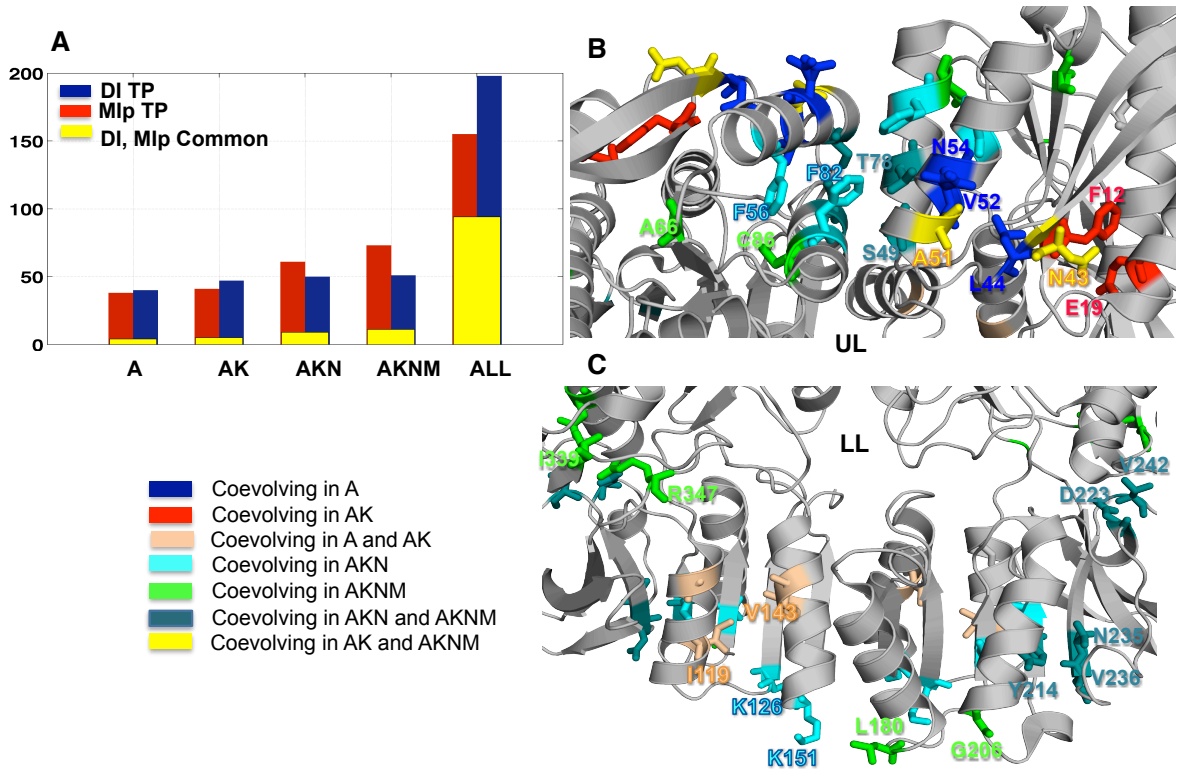


Figure 52. Most robust signals identified from MIp and DI mapped onto GluA2 structure.

(A) Shows the number of TPs (structural contacts that have sequence separation >5 and Cα distance < 10A) in each of the subsets for MIp and DI. The number of TPs that are identified by both methods are shown in the overlapping yellow bar. TPs that are identified as coevolving by both methods (Table 1, diagonal elements, numbers in parenthesis, except in set *ALL*) are mapped onto the structure shown in (B) for upper lobe (UL) and (C) for lower lobe (LL) of GluA2 (PDB ID: 3HSY), color coded as shown in bottom left panel.

The much larger set of common TPs predicted by both DI and MIp are shown in the network representation form in **Figure 53A.** We see strong interconnected webs of coevolutionary pairs, with residues that cluster together are colored based on their cluster. These residues are mapped on the structure (see colors) with residues that are hubs (most no of correlations) highlighted. This network of residues shows a possible mode of information propagation. Starting from the UL interfacial region (*orange*), we see the signal dissipate through neighboring regions and the cleft (*yellow*/ *red*), towards the LL (*light* and *dark green*) and also to

the flexible αF helix (*cyan*). This closely resembles the mode of information propagation as seen in allosteric signaling via the NTD in the AMPA receptor (see section 4.). The effect of double mutations on the pairs will help shed light on the importance of these residues in functioning of the iGluR receptors.

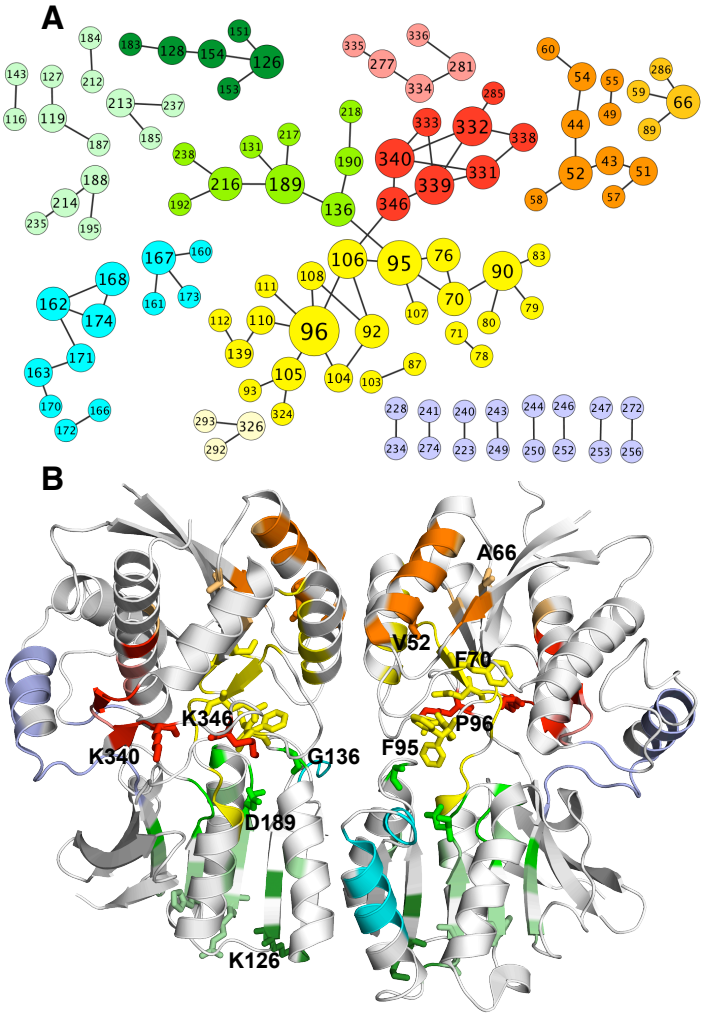


Figure 53. Network of coevolving residues mapped onto GluA2 structure.

(**A**) The most robust signals identified by both methods for the sequence class obtained from Pfam (Table 1, last diagonal element), are plotted as a network such that there exists an edge between nodes (residues) if they are coevolving. There are a total of 114 residues forming 94 edges (94 pairs identified by both DI and MIp), with the node size being directly proportional to the no. of edges for each node. Those that coevolve with a larger number of other residues are hubs that may be particularly important for function/signaling. They are colored according to how these residues cluster on the structure as shown in (**B**), with a few hubs shown explicitly.

Another noticeable albeit intuitive observation from this study is that residues that are conserved in smaller subsets are in fact those that coevolve together or with others in larger subsets (coevolving here implies those that are common TPs between MIp and DI) (**Figure 54**). This highlights intricacies in natural evolution, where residues that are important for structure and function (as suggested by virtue of high conservation in smaller subfamilies) are allowed to undergo specific changes that are compensated by neighboring mutations. This maintains the overall integrity of the structure while allowing for subtle changes that facilitate distinct behaviors seen in subtype specific iGluRs. Thus, overall we uncover important evolutionary dynamical behavior of the iGluR NTDs and shed light on current methods for coevloutionary analysis, their advantages and drawbacks, which can be further explored by larger-scale analysis of all existing protein families.

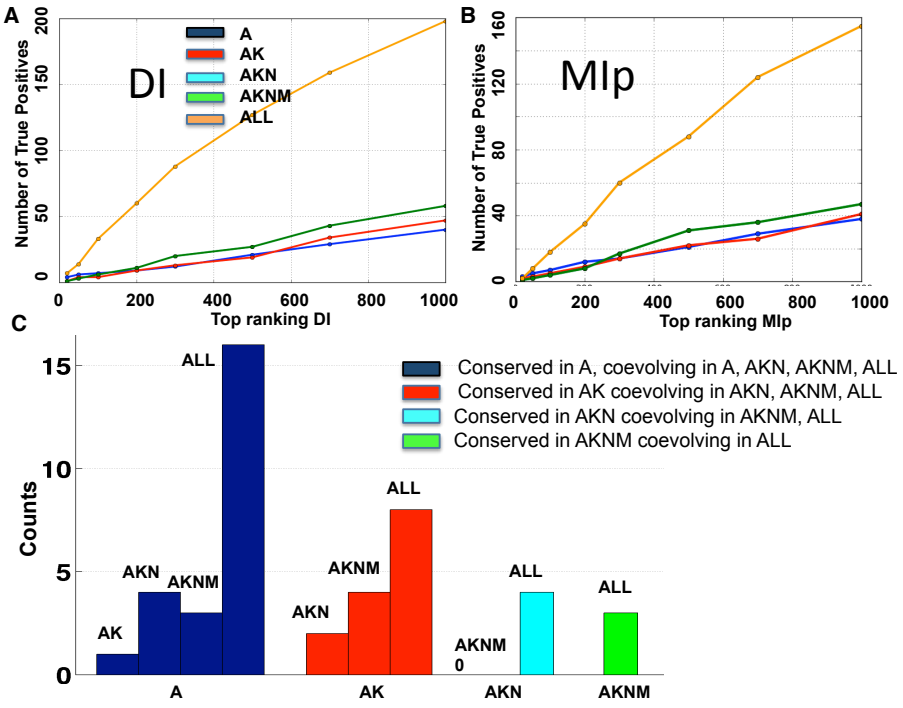


Figure 54. Coevolution properties of iGluR family members (including substructures) obtained from hierarchical analysis.

(A) Shows the cumulative distribution of TPs within the 1000 top ranking pairs identified by DI for each set. Similar as in **(A)** for MIp (color similar as **A**). **(C)** Shows the bar plot for the noumber of conserved residues in smaller subsets that become coevolving in larger ones. Each bar within a group eg: *A* (blue), represents the larger groups (in this case 4, *AK, AKN, AKNM* and *ALL*), that shows the no. of conserved residues in *A* that are coevolving in A, AK. AKNM and ALL (each of the four *blue* bars).

6.4 CONCLUSION

The present study describes the development of new software for a rapid assessment and visualization of sequence conservation and coevolution patterns in relation to structure-encoded dynamics of proteins, motivated by our current understanding of the significant role of intrinsic dynamics in the evolutionary selection of sequence. To enable an efficient, insightful, and integrated study of sequence-structure-dynamics features, we chose to build the new computing software as an extension of *ProDy* (Bakan et al., 2011), which enables rapid evaluation of spatial movements and cross-correlations favored by the query protein structure using both experimental and computational data. *ProDy* API and new extensions integrate well with open-source packages NumPy and Matplotlib to harness their efficient and powerful features. This makes the API suitable for interactive usage and enables rapid and easy development of new applications. These aspects of *ProDy* have been proven useful, as demonstrated by its integration to a set of diverse servers and applications (Andrei et al., 2012; Collier et al., 2012; Ho et al., 2012; Perez et al., 2012). We view the current development as a step toward gaining a better understanding of proteins' functional mechanisms and their evolution using sequence, structure and dynamics information. The significance of this development is also depicted in facilitating large-scale studies of protein families and identifying meaningful associations between evolutionary patterns, as evidenced here in the application to the GluR family. We anticipate further

development of this scalable resource and suitable integration with other tools and databases that provide complementary data, driven by the feedback from the community of researchers.

The development of "*evol*" was done in collaboration with Dr. Ahmet Bakan, and the hierarchical analysis work was done in collaboration with Dr. Ingo Greger. The results from this chapter are being organized into a manuscript: Dutta A*, Bakan A*, and Bahar I. *Evol: Comparative Analysis of Protein Sequence Co-evolution and Structural Dynamics.*

7.0 CONCLUSION AND FUTURE WORK

In this dissertation we have performed a comprehensive analysis of the iGluR family, using a combination of approaches: coarse-grained ENM analysis, all-atom MD simulations, graph network inference methods as well as sequence correlation analyses. Using NMA-ENM, we gain a deeper understanding of the conformational space accessible to iGluR NTDs. The advantage of NMA-ENM analyses is that they can be performed very fast on individual structures and do not depend on the availability of numerous family representatives (that are needed for principal components analyses, for example). This gives us an opportunity to study the dynamics of the intact receptor of AMPAR (the only one crystallized so far) whose dynamics may be beyond reach of conventional MD simulations (especially at milliseconds timescale). Modeling proteins using network models also enables us to decipher a pathway of communication via residues that appear as "hubs" of information flow identified from topological characteristics, informational theoretic and network perturbation approaches. In a sense, graph/network representation allows us to bridge between structural dynamics and allosteric signaling. In either case, the native contact topology defines the most cooperative modes of motions, which, in turn, define efficient communication mechanisms, or the potential structural changes in different forms or environments.

Our studies open a wide avenue for testing the functional significance of dynamics and allosteric potential in iGluR, specifically AMPARs. Experimental corroborations of ENM-

predicted dynamics in AMPARs (as seen for NMDARs in Zhu et al., 2013) might be obtained through cysteine mutational and cross-linking experiments that may suggest the existence of alternate conformations sampled by the receptor during its function. Experiments studying the functional impact of mutations in residues predicted as having high allosteric potential will throw some light on behavioral characteristics of AMPARs and at the same time illustrate the strength of our predictions.

The importance of incorporating receptor and ligand flexibility in identifying binding sites in proteins has been widely recognized. MD simulations of iGluRs in the presence of organic probes that resemble drug-like fragments enable us to capture protein flexibility and solvation effects. Using this method, we decipher important binding characteristics of iGluR NTD dimers and monomers. A primary highlight from this study (Chapter 5) is the identification of a novel binding site in GluA3 dimer LL interface. Here, we present a pharmacophore model that captures the essential features of this binding site, and opens a lucrative opportunity to identify compounds that may bind to this site. Virtual screening with compounds from existing databases in the first pass can help identify a few compounds that conform to the chemical features, shape and volume of this site. This may help in refining the binding site pharmacophore model increasing the probability of identifying compounds (through a second iteration of virtual screening) that may bind GluA3 with higher efficacy. The next step would be to test the top predictions experimentally to ascertain their effect on receptor function. Thus our findings pave the way for discovery of drugs/compounds that regulate the gating dynamics of iGluRs.

The correlation between sequence and structural dynamics cannot be ignored. With the growing repertoire of sequences (from fast sequencing techniques), it becomes essential to understand the relationship between sequence-structure-function so that we can extract

meaningful results from sequence analysis that can offer functional insights about the protein even in the absence of structural data. Towards this end, we developed "*evol*" for fast sequence-conservation/co-evolution analysis. Combined with *ProDy* (protein dynamics analysis package), this offers a unique platform for sequence-structural dynamics comparisons by the broader scientific community. There is a large potential for growth of this software package. As algorithms for examining sequence correlations evolve, these newer methods can be incorporated into *evol* so that users have a larger and better repertoire of methods to choose from. Large-scale analyses of sequences available for all protein-families as well as their structural counterparts offers another new avenue where our contribution can be useful in extracting information that can be used to fill the gaps in the sequence-structure space in the absence of sufficiently broad structural data for a given family.

APPENDIX A

APPENDIX A: METAL BINDING PROTEIN DATASETS

Dataset I of 30 metal-binding proteins structurally resolved in both *apo* and *holo* forms.

HOLO ID and CHAIN		LENGTH	APO PDB ID and CHAIN		LENGTH	MET-AL	METAL BINDING RESIDUES	BACKBONE RMSD	METAL-BINDING SITE RMSD [(a)]	GLOBAL dynamics CORREL[(b)]	Global Dynamics with Metal	B-Factor CORREL
1eu4	A	204	1et9	A	195	Zn	D160,H198,D200	0.189	0.155	0.967	0.972	0.912
1xll	A	393	1xla	A	393	Zn	H219,E216,D254, D256 E180,D244,D292	0.152	0.254	1.000	1.000	0.949
1l0y	B	220	1fnu	A	221	Zn	H10,E189	0.362	0.003	0.970	0.966	0.506
1mfm	A	153	10zt	M	153	Zn	H63,H71,H80,D83	0.527	0.274	0.987	0.970	0.206
1i6n	A	278	1i60	A	278	Zn	E142,E246,H200, D174	0.095	0.168	0.998	0.998	0.932
1lt8	A	348	1lt7	A	305	Zn	C217,C299,C300	0.347	0.088	0.977	0.977	0.924
1toa	A	277	1kof	A	277	Zn	H199,H68,H133, D279	0.624	0.055	0.994	0.993	0.456
1i0d	A	331	1pta	--	318	Zn	H55,H57,D301	1.182	1.740	0.719	0.729	0.514
1eu3	A	210	1et6	A	201	Zn	D204,H162,H202	0.159	0.108	0.985	0.984	0.907
1moo	A	256	2cbe	--	258	Zn	H94,H96,H119	0.194	0.029	0.988	0.974	0.893
1im5	A	179	1ilw	A	179	Zn	H54,D52,H71	0.163	0.079	0.996	0.912	0.880
1e67	A	128	1e65	--	128	Zn	H117,G45,H46, C112, M121	0.180	0.448	0.998	0.992	0.900
1c3r	A	372	1c3p	A	372	Zn	H170,D168,D258	0.280	0.068	0.952	0.951	0.826
1c24	A	262	4mat	A	271	Co	E235,E204,D108, H171, D97	0.416	0.152	0.895	0.884	0.651
1vlx	A	128	1e65	A	128	Co	H117,H46,C112	0.246	0.366	0.978	0.984	0.799
1k1e	A	177	1j8d	A	180	Co	D14,D16,D107	0.341	0.148	0.994	0.998	0.841
1xim	A	392	7xim	A	390	Co	E181,E217,D245, D292	0.065	0.061	1.000	1.000	0.962
1iid	A	422	1iic	A	422	Ni	H38,D37,H296	0.356	0.618	0.814	0.856	0.658
1b9m	**A**	**258**	**1h9s**	**A**	**138**	**Ni**	**D139,D148,H140, H146**	**1.384**	**2.340**	**0.201**	**0.321**	**0.428**
1gp5	A	346	1gp4	A	346	Fe	H288,H232,D234	0.332	0.061	0.717	0.684	0.937
1iej	**A**	**329**	**1tfa**	**A**	**329**	**Fe**	**Y92,H250,D60, Y191**	**5.580**	**3.106**	**0.627**	**0.631**	**0.744**
1ltz	A	257	1ltu	A	284	Fe	H138,H143,E184	0.183	0.366	0.863	0.869	0.938
1fr4	A	274	2cbe	A	274	Cu	H96,H94,H119	0.401	0.126	0.994	0.986	0.846
1ivv	A	620	1avk	A	258	Cu	H592,H431,H433	0.349	0.037	0.985	0.994	0.653
1gz4	A	551	1qr6	A	551	Mn	E255,D256,D279	1.380	0.178	0.942	0.954	0.587
1muc	A	360	1bkh	A	358	Mn	D198,D249,E224	0.158	0.044	0.994	0.993	0.885
1fui	B	591	1fui	F	591	Mn	E337,H528,D361	0.092	0.020	1.000	1.000	0.916
1jfz	A	148	1i4s	A	147	Mn	E40,D107,E110	0.412	0.080	0.928	0.811	0.547
1jlk	A	141	1i3c	A	144	Mn	D68,D70,D16	0.540	0.327	0.886	0.843	0.676
1dck	B	125	1dbw	A	125	Mn	R56,D11,D54	0.183	0.125	0.950	0.927	0.481
AVG		*291*			*274*			*0.389*	*0.221*	*0.945*	*0.936*	*0.745*

(a) calculated after optimal structural alignment, average RMSDs refer to all pairs of structures except for two, *highlighted in boldface, which exhibit RMSD > 2.0Å in the backbone and/or metal-binding site.*
(b) Correlation between the GNM softest modes (histogram of residue displacements in mode 1) calculated for the holo and apo forms. For the pair (1c24, 4mat) the correlation between respective modes 1 and 2 is given, as these were the counterparts of each other; and for (1k1e, 1j89), the correlation between the softest modes 2 is reported.

DATASET II of 57 *holo* metal-binding proteins[(a)] and correlations between computed and experimental B-factors

HOLO PDB ID	*CHAIN*	LENGTH	METAL	METAL BINDING RESIDUES	GNM *vs* B-factor CORRELATION (ONE CHAIN)[(c)]	GNM *vs* B-factor CORRELATION (ALL CHAINS)
1eu4	A	204	Zn	D160,H198,D200	0.659	0.659
1xll	A	393	Zn	H219,E216,D254,D256 E180,D244,D292	0.424	0.416
1l0y	B	220	Zn	H10,E189	0.514	0.869
1mfm	A	153	Zn	H63,H71,H80,D83	0.332	0.332
1i6n	A	278	Zn	E142,E246,H200,D174	0.506	0.506
1lt8	A	348	Zn	C217,C299,C300	0.290	0.497
1toa	A	277	Zn	H199,H68,H133,D279	0.670	0.694
1i0d	A	331	Zn	H55,H57,D301	0.635	0.676
1eu3	A	210	Zn	D204,H162,H202	0.650	0.709
1moo	A	256	Zn	H94,H96,H119	0.628	0.628
1im5	A	179	Zn	H54,D52,H71	0.500	0.500
1e67	A	128	Zn	H117,G45,H46,C112, M121	0.611	0.736
1c3r	A	372	Zn	H170,D168,D258	0.788	0.761
1c24	A	262	Co	E235,E204,D108,H171, D97	0.661	0.661
1vlx	A	128	Co	H117,H46,C112	0.508	0.649
1k1e	A	177	Co	D14,D16,D107	0.513	0.626
1xim	A	392	Co	E181,E217,D245,D292	0.203	0.741
1iid	A	422	Ni	H38,D37,H296	0.457	0.436
1b9m	A	258	Ni	D139,D148,H140,H146	0.028	0.430
1gp5	A	346	Fe	H288,H232,D234	0.565	0.565
1iej	A	329	Fe	Y92,H250,D60,Y191	0.776	0.776
1ltz	A	257	Fe	H138,H143,E184	0.495	0.495
1fr4	A	274	Cu	H96,H94,H119	0.560	0.558
1ivv	A	620	Cu	H592,H431,H433	0.396	0.731
1gz4	A	551	Mn	E255,D256,D279	0.245	0.780
1muc	A	360	Mn	D198,D249,E224	0.489	0.461
1fui	B	591	Mn	E337,H528,D361	0.491	0.591
1jfz	A	148	Mn	E40,D107,E110	0.683	0.585
1jlk	A	141	Mn	D68,D70,D16	0.563	0.651
1dck	B	125	Mn	R56,D11,D54	0.677	0.677
1d09	A,B[(b)]	153	Zn	C109,C114,C141,C138	0.438	0.669
1enq	A	230	Zn	E8,D10,H24	0.639	0.583
1hp7	A	376	Zn	H73,H93,E89	0.701	0.701
1ec5	A	48	Zn	H38,E36,E10	0.205	0.423
1frp	A	320	Zn	E280,D118,D121	0.534	0.806
1b66	A	138	Zn	H48,H23,H50	0.656	0.677
1hr6	A	457	Zn	H70,H74,D150	0.622	0.727
1li5	A	386	Zn	H234,E238,C28,C209	0.642	0.642
1h0o	A	288	Co	H270,D233,D267,D170	0.626	0.626
1a0c	A	437	Co	E267,E231,D338,D295	0.449	0.779
1g2a	A	164	Ni	H132,H136,C90	0.676	0.519
1nzr	A	128	Ni	H46,C112,H117	0.391	0.607
1gy9	A	279	Fe	H255,H99,D101	0.373	0.373
1dqi	A	124	Fe	H47,H41,C111,H16,H114,E14	0.379	0.602
1bou	B	298	Fe	E242,H12,H61	0.528	0.587
1bsm	A	201	Fe	D161,H165,H75,H27	0.348	0.434
1mty	D	512	Fe	E243,E209,E144,H246	0.499	0.463
1dq6	A	237	Mn	D19,E8,D10,H24	0.697	0.697
1d3v	A	308	Mn	D128,D232,D124,H101	0.350	0.350
1fi2	A	201	Mn	H137,H88,H90,E95	0.406	0.406
1kgp	D	296	Mn	E108,E202,D77,H111	0.463	0.463
1g5b	C	221	Mn	D20,D49,H22,H186,H139	0.602	0.606
1ii7	A	333	Mn	H10,H208,D8,D49,H173,H206	0.575	0.739
1hfu	A	500	Cu	H457,C452,H396	0.636	0.636
1iby	A	112	Cu	H98,C95,H103,E60	0.737	0.282
1e30	A	153	Cu	H143,H85,C138,Q148	0.585	0.788
1juh	A	334	Cu	H68,H66,H112,E73	0.579	0.610
AVG					***0.53***	***0.61***

[(a)]The set comprises 30 holo proteins listed in Table S1 + 27 structures resolved in the holo form only.

[(b)]The functional unit for 1D09 includes both chain A, B and has been used in all further calculations.

[(c)]GNM calculations are performed for a single chain, in the absence of neighboring chains in the PDB entry. The results in the last column are obtained with GNM calculations performed for all chains in a given PDB entry

DATASET III of 88 additional metal-binding proteins resolved in *holo* form.

HOLO PDB ID	CHAIN	METAL	METAL BINDING RESIDUES	HOLO PDB ID	CHAIN	METAL	METAL BINDING RESIDUES
2bc2	A	Zn	H86,H88,H149	1jw9	B	Zn	C172,C175,C244,C247
1j7n	A	Zn	H690,H686,E735	1vhh	A	Zn	H141,H183,D148
1g12	A	Zn	H121,H117,D130	1gpc	A	Zn	C90,C87,H64,C77
1dsz	A	Zn	C1177,C1290,C1171,C1187	1a5t	A	Zn	C59,C62,C65,C50
2a0b	A	Zn	E728,E754,E758	1a4m	A	Zn	H1515,H1517,H1714,D1795
1lml	A	Zn	H264,H334,H268	4enl	A	Zn	E295,D246,D320
1de5	A	Zn	E234,D267,D334,H294	2dtr	A	Co	H79,H98,E83
1h3n	A	Zn	C487,C484,C439,C442	4ubp	C	Ni	H275,H249,H139,H137,D363
1k3x	A	Zn	C240,C237,C257,C260	2tdx	A	Ni	H79,H98,E83,D102,E105,M10,H106
1h7n	A	Zn	C133,C135,C143	1eje	A	Ni	H68,E105,ASN42
1ld8	B	Zn	C799,H862,D797	1mrp	A	Fe	Y195,Y196,E57,H9
1j98	A	Zn	H58,H54,C126	1bk0	A	Fe	H214,D216,H270
1fn9	A	Zn	H71,C73,C54,C51	1han	A	Fe	H210,H146,E260,H189
1jpu	A	Zn	H274,D173,H256	3pcd	M	Fe	H460,Y408,H462
1hxr	A	Zn	C26,C23,C94,C97	1yge	A	Fe	H499,H690,H504,ILE839
1dcq	A	Zn	C267,C287,C284,C264	1dmh	A	Fe	H226,H224,Y164
1lr5	A	Zn	E63,H57,H106,H59	1rxf	A	Fe	H243,H183,D185
1ekm	A	Zn	H624,H456,H458,Y405	1brf	A	Fe	C5,C8,C38,C41
1a8h	A	Zn	C144,C127,H147,C130	1lcf	A	Cu	Y192,D60,Y92,H253,D395,H597,Y528
1kfi	A	Zn	D312,D308,D310	1iaa	A	Cu	H102,H96,H92,Y149
1e4c	P	Zn	H94,H92,H155,E73	1oxy	A	Cu	H364,H328,H324,H204,H173,H177
1dyq	A	Zn	H187,D227,H225	1e9p	A	Cu	H46,H44,H118
1j8f	A	Zn	C221,C224,C195,C200	1phm	A	Cu	H107,H108,H172,H242,H244,M314,H235
1evx	A	Zn	C125,C132,H134,C138	1plc	A	Cu	H87,H37,C84,M92
1bkc	E	Zn	H405,H415,H409	1ips	A	Mn	D216,H270,GLN330
1kjz	A	Zn	C60,C72,C63,C75	1e9g	B	Mn	D120,D152,D115,E58,D147,D152
1evl	A	Zn	H511,C334,H385	1rzd	A	Mn	H94,H96,H119
1pmi	A	Zn	E138,H113,H285,Q111	2mnr	A	Mn	E247,E221,D195
1irn	A	Zn	C42,C39,C9,C6	1ef2	A	Mn	H1246,H1272,H1134,H1136,D1360
1ptq	A	Zn	H231,C280,C264,C261	1lby	A	Mn	D85,D200,D82,E67,Leu84,D38,THR40
2hrv	A	Zn	C112,H114,C52,C54	1jlm	A	Mn	D242,SER142,SER144
1lba	A	Zn	H17,H122,C130	1igv	A	Mn	ASN56,D58,D54,E60
1ctt	A	Zn	C129,C132,H102	3tmy	A	Mn	D54,D10,Thr56
1gl4	A	Zn	H515,D511,H513	1g15	A	Mn	D10,D70,E18
1ast	A	Zn	Y149,H92,H102,H96	1hpu	A	Mn	H252,D84,H217,ASN116,GLN254,H43,D41
1bwn	A	Zn	C154,C155,C165,H143	4xis	A	Mn	D255,H220,E217,D257
1vsr	A	Zn	C66,C117,C73,H71	1d8h	B	Mn	E496,E305,E307
1lbu	A	Zn	H197,H154,D161	1pdz	A	Mn	E294,D319,D244
1ia9	A	Zn	H1751,C1814,H1808,C1810	2pal	A	Mn	E101,D90,D92,D94,E59,E62,D53,D51, SER55,F57
1epw	A	Zn	H229,H233,E267	1dah	A	Mn	D54,E115,Thr16
1i6p	A	Zn	C101,D44,H98,C42	1a6q	A	Mn	D60,D282,D239,Gly61
1eb6	A	Zn	H132,H128,D143	1f52	A	Mn	E220,E131,E212,E357,H269
1hc7	A	Zn	C458,C461,C427,C432	1gq6	A	Mn	H121,D235,D144,D148,D237,H146
1dy1	A	Zn	D207,H132,H134,H142	1ll2	A	Mn	H212,D102,D104

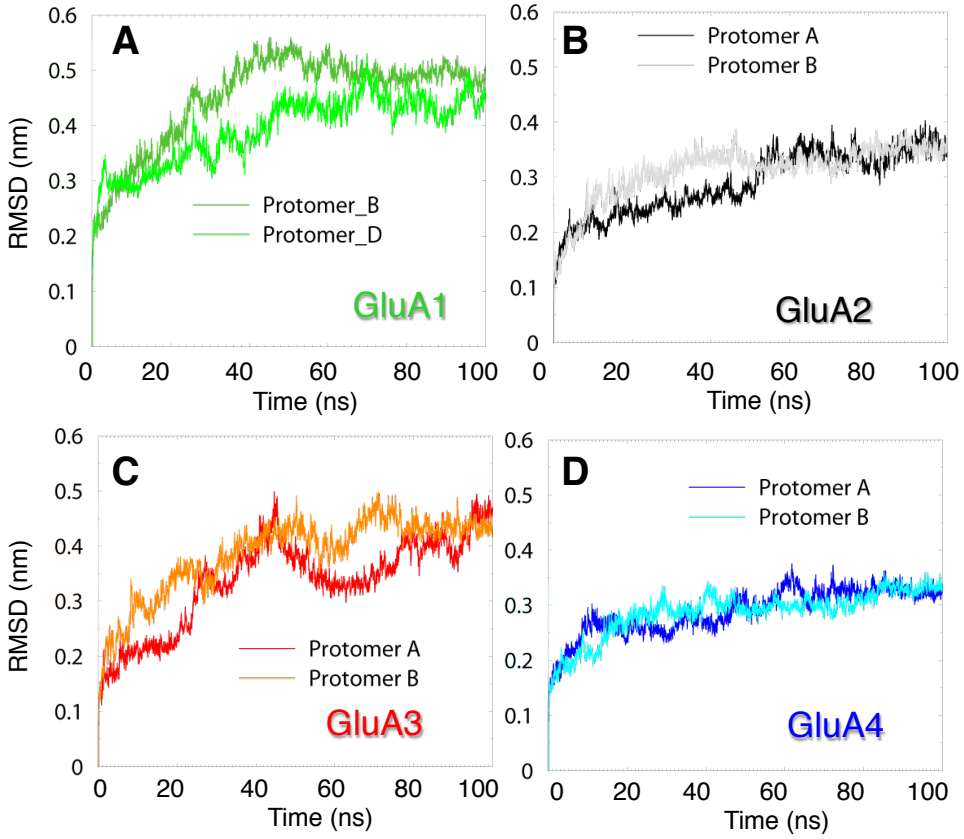
APPENDIX B

APPENDIX B: DETAILS OF MD SIMULATIONS

Structure	PDB-ID	Resolution	No. of Atoms	Simulation Time[a]
NMDA1	3JPW	2.80Å	78,095	50ns (0.47)
NMDA2	3JPY	3.21Å	78,066	50ns (0.35)
GluA1-*BD*	3SAJ	2.5Å	76,167	100ns (0.55)
GluA2	3HSY	1.75Å	75,262	100ns (0.35)
GluA3 -*CD*	3O21	2.2Å	76,345	100ns (0.47)
GluA4	Unpublished		76,231	100ns (0.35)
L144D	Model		76,257	100ns (0.45)
R163I	Model		76,324	100ns (0.45)

MD Simulations. The GROMACS program (Van Der Spoel et al., 2005) was used to generate MD trajectories for the systems listed in above. The proteins were solvated with single point charge (Berendsen et al., 1981) water molecules, using GROMOS 43a1 force field (Lindahl et al., 2001). MD runs were performed at 310K (by implementing Berendsen's temperature coupling to protein and water molecules) and atmospheric pressure. Electroneutrality was achieved by adding counterions. Electrostatic interactions were treated with the particle mesh Ewald method (Darden and Pedersen, 1993) and the LINCS (Hess et al., 1997) algorithm was used to constrain the bond lengths, enabling an integration timestep of 2fs. Each system was

energy minimized using the steepest descent algorithm, followed by an equilibration of 2ns, before the productive runs of 100 ns. During equilibration, backbone atoms were restrained by harmonic potentials, while side-chain atoms and water molecules were allowed to relax. The following figure displays the RMSDs from the initial state averaged over all residues as a function of time for the four AMPAR NTD subtypes.

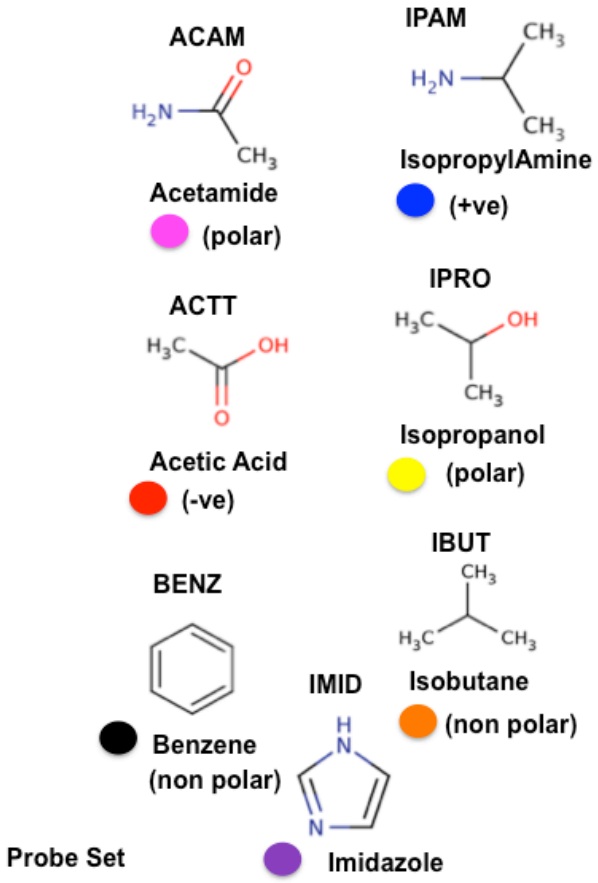


Comparison of the time evolution of RMSDs in residue positions for AMPAR NTDs. Results are shown for both protomers in (**A**) GluA1 (**B**) GluA2 and (**C**) GluA3, and (**D**) GluA4. A departure of ~0.35 nm from the starting structure is observed in GluA2 and GluA4 protomers. In GluA3 this value reaches ~0.47nm (originating from the enhanced mobility of the LL helices αE and αF, in particular) and in GluA1, 0.55nm (due to large mobility of the αH region).

APPENDIX C

APPENDIX C: DETAILS FOR DRUGGABILITY SIMULATIONS

System			PDB ID	Time (ns)	Probe Composition	Water	System Size
NTD Monomers	AMPA	GluA1	3SAJ - *C*	40	185 IPRO, 37 ACAM, 37 ACTT, 37 IPAM, 37 IBUT, 37 IMID (370 Total)	7400	32,403
		GluA2	3HSY - *B*	40	180 IPRO, 36 ACAM, 36 ACTT, 36 IPAM, 36 IBUT, 36 IMID (360 Total)	7200	31,612
		GluA3[(a)]	3O21 - *C*	40	224 IPRO, 32 ACAM, 32 ACTT, 32 IPAM (320 Total)	6400	28,974
		GluA4	4GPA - *A*	40	175 IPRO, 35 ACAM, 35 ACTT, 35 IPAM, 35 IBUT, 35 IMID (350 Total)	7000	31,016
	Kainate	GluK2	3H6G - *A*	40	170 IPRO, 34 ACAM, 34 ACTT, 34 IPAM, 34 IBUT, 34 IMID (340 Total)	6800	30,352
		GluK3	3OLZ - *A*	40	170 IPRO, 34 ACAM, 34 ACTT, 34 IPAM, 34 IBUT, 34 IMID (340 Total)	6800	30,389
		GluK5	3OM1 - *B*	40	170 IPRO, 34 ACAM, 34 ACTT, 34 IPAM, 34 IBUT, 34 IMID (340 Total)	6800	30,235
	NMDA	NR2B (1)	3JPY - *A*	20	259 IPRO, 37 ACAM, 37 ACTT, 37 IPAM (370 Total)	7400	32,201
		NR2B (2)[(b)]	3JPY - *A*	40	58 IPRO, 29 ACAM, 29 ACTT, 29 IBUT, 29 IMID, 87 BENZ (290 Total)	5800	26,460
		NR1	3Q41 - *A*	40	145 IPRO, 29 ACAM, 29 ACTT, 29 IPAM, 29 IBUT, 29 IMID (290 Total)	5800	26, 551
NTD Dimers	AMPA	GluA1	3SAJ - *BD*	60	336 IPRO, 84 ACAM, 126 ACTT, 126 IPAM, 126 IBUT, 42 BENZ (840 Total)	16,800	72,146
		GluA2	3HSY - *AB*	60	304 IPRO, 76 ACAM, 114 ACTT, 114 IPAM, 114 IBUT, 38 BENZ (760 Total)	15,200	66,332
		GluA3	3O21 - *CD*	60	228 IPRO, 72 ACAM, 108 ACTT, 108 IPAM, 108 IBUT, 36 BENZ (720 Total)	14,400	63,768
	NMDA	N2B-N1	3QEL - *AB*	60	294 IPRO, 84 ACAM, 126 ACTT, 126 IPAM, 126 IBUT, 84 BENZ (840 Total)	16,800	71,395
		N2B-N1 Mode 2	3QEL - *AB*	60	304 IPRO, 76 ACAM, 114 ACTT, 114 IPAM, 114 IBUT, 38 BENZ (760 Total)	15,200	66,007
LBD	M	A2 Monomer	1FTO - *A*	40	130 IPRO, 26 ACAM, 26 ACTT, 26 IPAM, 26 IBUT, 26 IMID (260 Total)	5200	22,607
	D	A2 Dimer (1)	1FTO - *AB*	40	215 IPRO, 43 ACAM, 43 ACTT, 43 IPAM, 43 IBUT, 43 IMID (430 Total)	8600	38,797
		A2 Dimer (2)	1FTO - *AB*	40	88 IPRO, 44 ACAM, 108* ACTT, 108 IPAM, 44 IBUT, 44 IMID (430 Total)	8800	39,312



Shows the structure of the probe fragments used in various proportions for Druggability simulations. The color-coded spheres indicate the colors that are used to represent the center of mass of probes in **Figures 39-47**.

APPENDIX D

APPENDIX D: MSAS FOR CONSERVATION ANALYSIS

Pfam alignments for conservation analysis		
	NTD	**LBD**
NTD of AMPA Receptor (PDB 3KG2)	10 – 384	393 – 506, 632 – 783
Pfam Family	PF1094	PF00497
Total no. of sequences	3297 sequences (dated March 12th 2012)	20060 sequences (dated March 29th 2012)
Length of MSA	2148	1870
Uniprot ID associated with PDB	GRIA2_RAT	GRIA1_RAT
Residues of PDB resolved in MSA	32-364 (w.r.t 3KG2)	411 – 783 for GluA1 (LBD and TMD)
Columns after label refinement (1)	Label = GRIA2_RAT 332	Label = GRIA1_RAT 373 (246 after removing TMD region) *
No of Sequences after row occupancy (at least 80%) and sequence identity refinement (at most 98%) (2)	1468	599

*Removing regions of TMD in the LBD alignment, the MSA columns retained for LBD were 411-520, 646-783, 246 residues. This was then aligned to the 3KG2 LBD, which corresponds to LBD residues 397-506, 632-769.

(1) The first refinement step involves removal or slicing of the columns in the MSA such that all gaps for a given sequence (corresponding to the representative PDB structure) can be eliminated. This step is necessary because we focus on the portions of the MSA that have structurally resolved counterparts in the PDB. It also reduces the length of the MSA to the length of the PDB resolved reference structure.

(2) This is followed by removal of sequences with >20% gaps and sequences with similarity >98% thereby reducing the number of sequences. This step is necessary for reducing redundancy and highly gapped sequences that add noise in calculations.

BIBLIOGRAPHY

Ahmed, A.H., et al. (2007) Dynamics of the S1S2 glutamate binding domain of GluR2 measured using 19F NMR spectroscopy, J Biol Chem, 282, 12773-12784.

Ahmed, A.H., et al. (2013) Dynamics of Cleft Closure of the GluA2 Ligand Binding Domain in the Presence of Full and Partial Agonists Revealed by Hydrogen-Deuterium Exchange, J Biol Chem.

Allen, K.N., et al. (1996) Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins, The Journal of Physical Chemistry, 100, 2605-2611.

Andrei, R.M., et al. (2012) Intuitive representation of surface properties of biomolecules using BioBlender, BMC Bioinformatics, 13 Suppl 4, S16.

Arinaminpathy, Y., Sansom, M.S. and Biggin, P.C. (2002) Molecular dynamics simulations of the ligand-binding domain of the ionotropic glutamate receptor GluR2, Biophys J, 82, 676-683.

Armstrong, N. and Gouaux, E. (2000) Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core, Neuron, 28, 165-181.

Armstrong, N., et al. (2006) Measurement of conformational changes accompanying desensitization in an ionotropic glutamate receptor, Cell, 127, 85-97.

Ashkenazy, H., et al. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, Nucleic Acids Res, 38, W529-533.

Atchley, W.R., et al. (2000) Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis, Mol Biol Evol, 17, 164-178.

Atilgan, A.R., Akan, P. and Baysal, C. (2004) Small-world communication of residues and significance for protein dynamics, Biophys J, 86, 85-91.

Atilgan, A.R., et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model, Biophys J, 80, 505-515.

Atilgan, C. and Atilgan, A.R. (2009) Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein, PLoS Comput Biol, 5, e1000544.

Auld, D.S. (2001) Zinc coordination sphere in biochemical zinc sites, Biometals, 14, 271-313.

Ayalon, G., et al. (2005) Two regions in the N-terminal domain of ionotropic glutamate receptor 3 form the subunit oligomerization interfaces that control subtype-specific receptor assembly, J Biol Chem, 280, 15053-15060.

Ayalon, G. and Stern-Bach, Y. (2001) Functional assembly of AMPA and kainate receptors is mediated by several discrete protein-protein interactions, Neuron, 31, 103-113.

Babor, M., et al. (2008) Prediction of transition metal-binding sites from apo protein structures, Proteins, 70, 208-217.

Bahar, I., et al. (1998) Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability, Phys. Rev. Lett., 80, 2733-2736.

Bahar, I., Atilgan, A.R. and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Fold Des, 2, 173-181.

Bahar, I., Chennubhotla, C. and Tobi, D. (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation, Curr Opin Struct Biol, 17, 633-640.

Bahar, I., et al. (2010a) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins, Chem Rev, 110, 1463-1497.

Bahar, I., et al. (2010b) Global dynamics of proteins: bridging between structure and function, Annu Rev Biophys, 39, 23-42.

Bahar, I. and Rader, A.J. (2005) Coarse-grained normal mode analysis in structural biology, Curr Opin Struct Biol, 15, 586-592.

Bakan, A. and Bahar, I. (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding, Proc Natl Acad Sci U S A, 106, 14349-14354.

Bakan, A., Meireles, L.M. and Bahar, I. (2011) ProDy: protein dynamics inferred from theory and experiments, Bioinformatics, 27, 1575-1577.

Bakan, A., et al. (2011) Improved assessment of druggability: exploring target protein dynamics in the presence of probe molecules, In Preparation.

Bakan, A., et al. (2012) Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules, Journal of chemical theory and computation, 8, 2435-2447.

Berman, H.M., et al. (2000) The Protein Data Bank, Nucleic Acids Res, 28, 235-242.

Bhabha, G., et al. (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis, Science, 332, 234-238.

Bigge, C.F. (1999) Ionotropic glutamate receptors, Curr Opin Chem Biol, 3, 441-447.

Borschel, W.F., et al. (2011) NMDA receptor activation requires remodelling of intersubunit contacts within ligand-binding heterodimers, Nature communications, 2, 498.

Bowie, D. (2008) Ionotropic glutamate receptors & CNS disorders, CNS Neurol Disord Drug Targets, 7, 129-143.

Brenke, R., et al. (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques, Bioinformatics, 25, 621-627.

Brooks, B. and Karplus, M. (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, Proc Natl Acad Sci U S A, 80, 6571-6575.

Brown, S.P. and Hajduk, P.J. (2006) Effects of conformational dynamics on predicted protein druggability, ChemMedChem, 1, 70-72.

Burger, L. and van Nimwegen, E. (2010) Disentangling direct from indirect co-evolution of residues in protein alignments, PLoS Comput Biol, 6, e1000633.

Burger, P.B., et al. (2012) Mapping the binding of GluN2B-selective N-methyl-D-aspartate receptor negative allosteric modulators, Mol Pharmacol, 82, 344-359.

Calhoun, J.R., et al. (2008) Solution NMR structure of a designed metalloprotein and complementary molecular dynamics refinement, Structure, 16, 210-215.

Castagnetto, J.M., et al. (2002) MDB: the Metalloprotein Database and Browser at The Scripps Research Institute, Nucleic Acids Res, 30, 379-382.

Cavasotto, C.N., Kovacs, J.A. and Abagyan, R.A. (2005) Representing receptor flexibility in ligand docking through relevant normal modes, Journal of the American Chemical Society, 127, 9632-9640.

Cheng, A.C., et al. (2007) Structure-based maximal affinity model predicts small-molecule druggability, Nat Biotechnol, 25, 71-75.

Cheng, Q., et al. (2002) A vibrational spectroscopic investigation of interactions of agonists with GluR0, a prokaryotic glutamate receptor, Biochemistry, 41, 1602-1608.

Chennubhotla, C. and Bahar, I. (2007) Signal propagation in proteins and relation to equilibrium fluctuations, PLoS Comput Biol, 3, 1716-1726.

Clayton, A., et al. (2009) Crystal structure of the GluR2 amino-terminal domain provides insights into the architecture and assembly of ionotropic glutamate receptors, J Mol Biol, 392, 1125-1132.

Cline, M.S., et al. (2002) Information-theoretic dissection of pairwise contact potentials, Proteins, 49, 7-14.

Collier, J.H., et al. (2012) Super: a web server to rapidly screen superposable oligopeptide fragments from the protein data bank, Nucleic Acids Res, 40, W334-339.

Cover, T.M. and Thomas, J.A. (2001) Elements of Information Theory. Wiley-Interscience.

Cui, Q. and Bahar, I. (2006) Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. Chapman & Hall/CRC, London, UK.

Darden, T.Y.D. and Pedersen, L. (1993) Particle Mesh Ewald: An N.log(N) method for Ewald sums in large systems., J Chem Phys, 98, 10089-10092.

Das, U., et al. (2010) Domain organization and function in GluK2 subtype kainate receptors, Proc Natl Acad Sci U S A, 107, 8463-8468.

del Sol, A., et al. (2006) Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families, Protein Sci, 15, 2120-2128.

del Sol, A., et al. (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins, Mol Syst Biol, 2, 2006 0019.

Demirel, M.C., et al. (1998) Identification of kinetically hot residues in proteins, Protein Sci, 7, 2522-2532.

Dingledine, R., et al. (1999) The glutamate receptor ion channels, Pharmacol Rev, 51, 7-61.

Dunn, S.D., Wahl, L.M. and Gloor, G.B. (2008) Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction, Bioinformatics, 24, 333-340.

Dutta, A. and Bahar, I. (2010) Metal-binding sites are designed to achieve optimal mechanical and signaling properties, Structure, 18, 1140-1148.

Dutta, A., et al. (2012) Comparative dynamics of NMDA- and AMPA-glutamate receptor N-terminal domains, Structure, 20, 1838-1849.

Ebert, J.C. and Altman, R.B. (2008) Robust recognition of zinc binding sites in proteins, Protein Sci, 17, 54-65.

Eisenmesser, E.Z., et al. (2005) Intrinsic dynamics of an enzyme underlies catalysis, Nature, 438, 117-121.

Erreger, K., et al. (2004) Glutamate receptor gating, Critical reviews in neurobiology, 16, 187-224.

Eyal, E., Dutta, A. and Bahar, I. (2011) Cooperative Dynamics of Proteins Unraveled by Network Models, Wiley Interdisciplinary Reviews: Computational Molecular Science, 1, 426-439.

Eyal, E., et al. (2007) A pair-to-pair amino acids substitution matrix and its applications for protein structure prediction, Proteins, 67, 142-153.

Eyal, E., et al. (2004) Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins, J Comput Chem, 25, 712-724.

Eyal, E., Yang, L.W. and Bahar, I. (2006) Anisotropic network model: systematic evaluation and a new web interface, Bioinformatics, 22, 2619-2627.

Farina, A.N., et al. (2011) Separation of domain contacts is required for heterotetrameric assembly of functional NMDA receptors, J Neurosci, 31, 3565-3579.

Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching, Nucleic Acids Res, 39, W29-37.

Floquet, N., et al. (2006) Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors, FEBS letters, 580, 5130-5136.

Flory, P. (1976) Statistical thermodynamics of random networks. , Proc. R. Soc. Lond. A 351, 351-380

Fodor, A.A. and Aldrich, R.W. (2004) Influence of conservation on calculations of amino acid covariance in multiple sequence alignments, Proteins, 56, 211-221.

Furukawa, H., et al. (2005) Subunit arrangement and function in NMDA receptors, Nature, 438, 185-192.

Geremia, S., et al. (2005) Response of a designed metalloprotein to changes in metal ion coordination, exogenous ligands, and active site volume determined by X-ray crystallography, Journal of the American Chemical Society, 127, 17266-17276.

Gielen, M., et al. (2008) Structural rearrangements of NR1/NR2A NMDA receptors during allosteric inhibition, Neuron, 57, 80-93.

Gielen, M., et al. (2009) Mechanism of differential control of NMDA receptor activity by NR2 subunits, Nature, 459, 703-707.

Glembo, T.J., et al. (2012) Collective dynamics differentiates functional divergence in protein evolution, PLoS Comput Biol, 8, e1002428.

Go, N., Noguti, T. and Nishikawa, T. (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes, Proc Natl Acad Sci U S A, 80, 3696-3700.

Golovin, A., et al. (2005) MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites, Proteins, 58, 190-199.

Gourley, D.G., et al. (2001) Oxyanion binding alters conformation and quaternary structure of the c-terminal domain of the transcriptional regulator mode. Implications for molybdate-dependent regulation, signaling, storage, and transport, J Biol Chem, 276, 20641-20647.

Greger, I.H., Ziff, E.B. and Penn, A.C. (2007) Molecular determinants of AMPA receptor subunit assembly, Trends Neurosci, 30, 407-416.

Guvench, O. and MacKerell, A.D., Jr. (2009) Computational fragment-based binding site identification by ligand competitive saturation, PLoS Comput Biol, 5, e1000435.

Hajduk, P.J., Huth, J.R. and Fesik, S.W. (2005) Druggability indices for protein targets derived from NMR-based screening data, J Med Chem, 48, 2518-2525.

Haliloglu, T. Bahar, I. and Erman, B. (1997) Gaussian Dynamics of Folded Proteins, Phys Rev Lett, 79, 3090-3093.

Haliloglu, T., Bahar, I. and Erman, B. (1997) Gaussian dynamics of folded proteins, Phys. Rev. Lett. , 79, 3090-3093.

Haliloglu, T., et al. (2005) How similar are protein folding and protein binding nuclei? Examination of vibrational motions of energy hot spots and conserved residues, Biophys J, 88, 1552-1559.

Hansen, K.B., Furukawa, H. and Traynelis, S.F. (2010) Control of assembly and function of glutamate receptors by the amino-terminal domain, Mol Pharmacol, 78, 535-549.

Hansen, K.B., Yuan, H. and Traynelis, S.F. (2007) Structural aspects of AMPA receptor activation, desensitization and deactivation, Curr Opin Neurobiol, 17, 281-288.

He, X.L., et al. (2005) A new paradigm for hormone recognition and allosteric receptor activation revealed from structural studies of NPR-C, Peptides, 26, 1035-1043.

Hinsen, K. (1998) Analysis of domain motions by approximate normal mode calculations, Proteins, 33, 417-429.

Hinsen, K., et al. (2000) Harmonicity in slow protein dynamics, Chemical Physics, 261, 25-37.

Ho, H.K., et al. (2012) BetaSearch: a new method for querying beta-residue motifs, BMC Res Notes, 5, 391.

Hogner, A., et al. (2002) Structural basis for AMPA receptor activation and ligand selectivity: crystal structures of five agonist complexes with the GluR2 ligand-binding core, J Mol Biol, 322, 93-109.

Holloway, D.E., et al. (2009) Influence of naturally-occurring 5'-pyrophosphate-linked substituents on the binding of adenylic inhibitors to ribonuclease a: an X-ray crystallographic study, Biopolymers, 91, 995-1008.

Hopf, T.A., et al. (2012) Three-dimensional structures of membrane proteins from genomic sequencing, Cell, 149, 1607-1621.

Hopkins, A.L. and Groom, C.R. (2002) The druggable genome, Nat Rev Drug Discov, 1, 727-730.

Horning, M.S. and Mayer, M.L. (2004) Regulation of AMPA receptor gating by ligand binding core dimers, Neuron, 41, 379-388.

Huang, N. and Jacobson, M.P. (2010) Binding-site assessment by virtual fragment screening, PLoS One, 5, e10109.

Hunter, J.D. (2007) Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95.

Isin, B., et al. (2008) Mechanism of signal propagation upon retinal isomerization: insights from molecular dynamics simulations of rhodopsin restrained by normal modes, Biophys J, 95, 789-803.

Ivetac, A. and McCammon, J.A. (2010) Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach, Chem Biol Drug Des, 76, 201-217.

Jain, E., et al. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website, BMC Bioinformatics, 10, 136.

James, L.C., Roversi, P. and Tawfik, D.S. (2003) Antibody multispecificity mediated by conformational diversity, Science, 299, 1362-1367.

Jensen, M.H., et al. (2011) Intrinsic motions in the N-terminal domain of an ionotropic glutamate receptor detected by fluorescence correlation spectroscopy, J Mol Biol, 414, 96-105.

Jin, R., et al. (2003) Structural basis for partial agonist action at ionotropic glutamate receptors, Nat Neurosci, 6, 803-810.

Jin, R. and Gouaux, E. (2003) Probing the function, conformational plasticity, and dimer-dimer contacts of the GluR2 ligand-binding core: studies of 5-substituted willardiines and GluR2 S1S2 in the crystal, Biochemistry, 42, 5201-5213.

Jin, R., et al. (2009) Crystal structure and association behaviour of the GluR2 amino-terminal domain, EMBO J, 28, 1812-1823.

Jones, D.T., et al. (2012) PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments, Bioinformatics, 28, 184-190.

Karakas, E., Simorowski, N. and Furukawa, H. (2009) Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit, EMBO J.

Karakas, E., Simorowski, N. and Furukawa, H. (2009) Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit, EMBO J, 28, 3910-3920.

Karakas, E., Simorowski, N. and Furukawa, H. (2011) Subunit arrangement and phenylethanolamine binding in GluN1/GluN2B NMDA receptors, Nature, 475, 249-253.

Kass, I. and Horovitz, A. (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations, Proteins, 48, 611-617.

Keller, T.H., Pichota, A. and Yin, Z. (2006) A practical view of 'druggability', Current opinion in chemical biology, 10, 357-361.

Kendrik, M.J., Plishka, M.J. and Robinson, K.D. (1992) Metals in Biological Systems. Ellis Horwood Series in Inorganic Chemistry. Prentice Hall Professional Technical Reference, New York.

Kumar, J. and Mayer, M.L. (2010) Crystal structures of the glutamate receptor ion channel GluK3 and GluK5 amino-terminal domains, J Mol Biol, 404, 680-696.

Kumar, J., et al. (2009) The N-terminal domain of GluR6-subtype glutamate receptor ion channels, Nat Struct Mol Biol, 16, 631-638.

Kunishima, N., et al. (2000) Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor, Nature, 407, 971-977.

Landau, M., et al. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures, Nucleic Acids Res, 33, W299-302.

Lange, O.F., et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution, Science, 320, 1471-1475.

Lau, A.Y. and Roux, B. (2007) The free energy landscapes governing conformational changes in a glutamate receptor ligand-binding domain, Structure, 15, 1203-1214.

Lau, A.Y. and Roux, B. (2011) The hidden energetics of ligand binding and activation in a glutamate receptor, Nat Struct Mol Biol, 18, 283-287.

Lee, C.H. and Gouaux, E. (2011) Amino terminal domains of the NMDA receptor are organized as local heterodimers, PLoS One, 6, e19180.

Levitt, M., Sander, C. and Stern, P.S. (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme, J Mol Biol, 181, 423-447.

Lin, C.T., et al. (2005) Protein metal binding residue prediction based on neural networks, Int J Neural Syst, 15, 71-84.

Lin, H.H., et al. (2006) Prediction of the functional class of metal-binding proteins from sequence derived physicochemical properties by support vector machine approach, BMC Bioinformatics, 7 Suppl 5, S13.

Lipinski, C.A., et al. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Advanced drug delivery reviews, 46, 3-26.

Liu, T., Whitten, S.T. and Hilser, V.J. (2007) Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble, Proc Natl Acad Sci U S A, 104, 4347-4352.

Liu, Y. and Bahar, I. (2012) Sequence evolution correlates with structural dynamics, Mol Biol Evol, 29, 2253-2263.

Liu, Y., Gierasch, L.M. and Bahar, I. (2010) Role of Hsp70 ATPase domain intrinsic dynamics and sequence evolution in enabling its functional interactions with NEFs, PLoS Comput Biol, 6.

Lockless, S.W. and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families, Science, 286, 295-299.

Ma, J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes, Structure, 13, 373-380.

MacKerell, A.D., et al. (2002) CHARMM: The Energy Function and Its Parameterization. Encyclopedia of Computational Chemistry. John Wiley & Sons, Ltd.

Madden, D.R. (2002) The structure and function of glutamate receptor ion channels, Nat Rev Neurosci, 3, 91-101.

Marks, D.S., et al. (2011) Protein 3D structure computed from evolutionary sequence variation, PLoS One, 6, e28766.

Marks, D.S., Hopf, T.A. and Sander, C. (2012) Protein structure prediction from sequence variation, Nature biotechnology, 30, 1072-1080.

Marsh, J.A., Teichmann, S.A. and Forman-Kay, J.D. (2012) Probing the diverse landscape of protein flexibility and binding, Curr Opin Struct Biol, 22, 643-650.

Martin, L.C., et al. (2005) Using information theory to search for co-evolving residues in proteins, Bioinformatics, 21, 4116-4124.

May, A. and Zacharias, M. (2008) Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking, J Med Chem, 51, 3499-3506.

Mayer, M.L. (2006) Glutamate receptors at atomic resolution, Nature, 440, 456-462.

McConkey, B.J., Sobolev, V. and Edelman, M. (2002) Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure, Bioinformatics, 18, 1365-1373.

McFeeters, R.L. and Oswald, R.E. (2002) Structural mobility of the extracellular ligand-binding core of an ionotropic glutamate receptor. Analysis of NMR relaxation dynamics, Biochemistry, 41, 10472-10481.

Micheletti, C. (2012) Comparing proteins by their internal dynamics: Exploring structure-function relationships beyond static structural alignments, Phys Life Rev.

Ming, D. and Wall, M.E. (2005) Allostery in a coarse-grained model of protein dynamics, Phys Rev Lett, 95, 198103.

Mony, L., et al. (2009) Structural basis of NR2B-selective antagonist recognition by N-methyl-D-aspartate receptors, Mol Pharmacol, 75, 60-74.

Morcos, F., et al. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families, Proc Natl Acad Sci U S A, 108, E1293-1301.

Morgan, D.H., et al. (2006) ET viewer: an application for predicting and visualizing functional sites in protein structures, Bioinformatics, 22, 2049-2050.

Nayal, M. and Honig, B. (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites, Proteins, 63, 892-906.

Nicolay, S. and Sanejouand, Y.H. (2006) Functional modes of proteins are among the most robust, Phys Rev Lett, 96, 078104.

Noivirt, O., Eisenstein, M. and Horovitz, A. (2005) Detection and reduction of evolutionary noise in correlated mutation analysis, Protein engineering, design & selection : PEDS, 18, 247-253.

O'Hara, P.J., et al. (1993) The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins, Neuron, 11, 41-52.

Pang, A., et al. (2005) Comparative molecular dynamics--similar folds and similar motions?, Proteins, 61, 809-822.

Passerini, A., et al. (2007) Predicting zinc binding at the proteome level, BMC Bioinformatics, 8, 39.

Passerini, A., et al. (2006) Identifying cysteines and histidines in transition-metal-binding sites using support vector machines and neural networks, Proteins, 65, 305-316.

Perez, A., et al. (2012) FlexE: Using Elastic Network Models to Compare Models of Protein Structure, Journal of chemical theory and computation, 8, 3985-3991.

Perin-Dureau, F., et al. (2002) Mapping the binding site of the neuroprotectant ifenprodil on NMDA receptors, J Neurosci, 22, 5955-5965.

Phillips, J.C., et al. (2005) Scalable molecular dynamics with NAMD, Journal of computational chemistry, 26, 1781-1802.

Pin, J.P., Galvez, T. and Prezeau, L. (2003) Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors, Pharmacol Ther, 98, 325-354.

Pohlsgaard, J., et al. (2011) Lessons from more than 80 structures of the GluA2 ligand-binding domain in complex with agonists, antagonists and allosteric modulators, Neuropharmacology, 60, 135-150.

Prim, R.C. (1957) Shortest connection networks and some generalizations, Bell System Technical Journal, 36, 1389-1401.

Procaccini, A., et al. (2011) Dissecting the specificity of protein-protein interaction in bacterial two-component signaling: orphans and crosstalks, PLoS One, 6, e19729.

Punta, M., et al. (2012) The Pfam protein families database, Nucleic Acids Res, 40, D290-301.

Quiocho, F.A. and Ledvina, P.S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes, Molecular microbiology, 20, 17-25.

Rambhadran, A., Gonzalez, J. and Jayaraman, V. (2010) Subunit arrangement in N-methyl-D-aspartate (NMDA) receptors, J Biol Chem, 285, 15296-15301.

Rossmann, M., et al. (2011) Subunit-selective N-terminal domain associations organize the formation of AMPA receptor heteromers, EMBO J, 30, 959-971.

Rossmann, M., et al. (2010) N-terminal domain associations organize the formation of AMPA receptor heteromers, submitted.

Salussolia, C.L., et al. (2011) Arrangement of subunits in functional NMDA receptors, J Neurosci, 31, 11295-11304.

Seco, J., Luque, F.J. and Barril, X. (2009) Binding site detection and druggability index from first principles, J Med Chem, 52, 2363-2371.

Sirrieh, R.E., Maclean, D.M. and Jayaraman, V. (2013) Amino-terminal Domain Tetramer Organization and Structural Effects of Zinc Binding in the N-Methyl-D-aspartate (NMDA) Receptor, J Biol Chem, 288, 22555-22564.

Smock, R.G. and Gierasch, L.M. (2009) Sending signals dynamically, Science, 324, 198-203.

Sobolevsky, A.I., Rosconi, M.P. and Gouaux, E. (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor, Nature, 462, 745-756.

Suel, G.M., et al. (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins (vol 10, pg 59, 2003), Nat Struct Biol, 10, 232-232.

Suhre, K. and Sanejouand, Y.H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement, Nucleic Acids Res, 32, W610-614.

Sukumaran, M., et al. (2011) Dynamics and allosteric potential of the AMPA receptor N-terminal domain, EMBO J, 30, 972-982.

Sun, Y., et al. (2002) Mechanism of glutamate receptor desensitization, Nature, 417, 245-253.

Tainer, J.A., Roberts, V.A. and Getzoff, E.D. (1992) Protein metal-binding sites, Curr Opin Biotechnol, 3, 378-387.

Tama, F. and Brooks, C.L. (2006) Symmetry, form, and shape: guiding principles for robustness in macromolecular machines, Annu Rev Biophys Biomol Struct, 35, 115-133.

Tillier, E.R. and Lui, T.W. (2003) Using multiple interdependency to separate functional from phylogenetic correlations in protein alignments, Bioinformatics, 19, 750-755.

Tirion, M.M. (1996) Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis, Phys Rev Lett, 77, 1905-1908.

Tobi, D. and Bahar, I. (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state, Proc Natl Acad Sci U S A, 102, 18908-18913.

Tokuriki, N. and Tawfik, D.S. (2009) Protein dynamism and evolvability, Science, 324, 203-207.

Trakhanov, S., et al. (2005) Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity, Biochemistry, 44, 6597-6608.

Traynelis, S.F., et al. (2010) Glutamate Receptor Ion Channels: Structure, Regulation, and Function, Pharmacol Rev, 62, 405-496.

Tsuchiya, D., et al. (2002) Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+, Proc Natl Acad Sci U S A, 99, 2660-2665.

Vanommeslaeghe, K., et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, Journal of computational chemistry, 31, 671-690.

Wainreb, G., et al. (2011) Protein stability: a single recorded mutation aids in predicting the effects of other mutations in the same amino acid site, Bioinformatics, 27, 3286-3292.

Waterhouse, A.M., et al. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, Bioinformatics, 25, 1189-1191.

Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks, Nature, 393, 440-442.

Weigt, M., et al. (2009) Identification of direct residue contacts in protein-protein interaction by message passing, Proc Natl Acad Sci U S A, 106, 67-72.

Wollenberg, K.R. and Atchley, W.R. (2000) Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap, Proc Natl Acad Sci U S A, 97, 3288-3291.

Worth, C.L., Gong, S. and Blundell, T.L. (2009) Structural and functional constraints in the evolution of protein families, Nat Rev Mol Cell Biol, 10, 709-720.

Yang, L.W. and Bahar, I. (2005) Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes, Structure, 13, 893-904.

Yao, G., et al. (2011) Crystal structure of the glutamate receptor GluA1 amino-terminal domain, Biochem J.

Zheng, W. and Brooks, B.R. (2005) Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin, Biophys J, 89, 167-178.

Zheng, W., et al. (2005) Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved, Structure, 13, 565-577.

Zhu, S., et al. (2013) Allosteric signaling and dynamics of the clamshell-like NMDA receptor GluN1 N-terminal domain, Nat Struct Mol Biol, 20, 477-485.